

Grado Universitario en Ingeniería Informática
2021-2022

Trabajo Fin de Grado

Preparación y configuración de un entorno de simulación de transporte urbano basado en agentes: Leganés en Matsim

Alejandro de la Parra Rebato

Tutor

Agapito Ismael Ledezma Espino

Leganés, Marzo 2022



Esta obra se encuentra sujeta a la licencia Creative Commons **Reconocimiento – No Comercial – Sin Obra Derivada**

Resumen

En un mundo cada vez más poblado y urbanizado, están surgiendo grandes cambios en la movilidad de la población, que se han acentuado tras la pandemia del Covid-19. La digitalización y la introducción de las tecnologías de la información y de la comunicación (TIC) y el Big Data, están permitiendo la aparición de nuevos medios de transporte y de formas de gestionar los desplazamientos en conjunto, para solucionar los problemas de movilidad en las ciudades.

Para estudiar la viabilidad de todas las ideas de mejora que están surgiendo, se utilizan simulaciones que permiten detectar problemas y estudiar sus soluciones antes de su implementación. Por ello, en este trabajo se quieren sentar las bases para las simulaciones de movilidad, exponiendo todos los procesos necesarios previos a la simulación. Uno de los cuales, es el sintetizador de población que se ha desarrollado en este trabajo para la generación de poblaciones sintéticas.

El sistema generado en este trabajo ha pretendido estudiar un enfoque distinto al habitual, creando hogares a partir de las estructuras sociales de la población objetivo, dejadas de lado por el resto de sintetizadores consultados. Para ello, se ha realizado un estudio sociológico y se han recopilado datos estadísticos de la población de Leganés, a partir de los cuales se crean las familias.

Dedicatoria

Después de tantos años estudiando durante las distintas etapas educativas, este proyecto supone el fin de esta etapa de mi vida. Una etapa que, pese a ciertas dificultades, ha estado llena de momentos felices en los que mis padres siempre me han acompañado. Quiero agradecer profundamente a mis padres todas las risas, enseñanzas, viajes, consuelos y apoyos tanto en los momentos buenos como en los malos. Sin vosotros no habría llegado nunca a ser la persona adulta que soy hoy. Gracias de todo corazón.

Querría agradecer a Adina su amor y su apoyo, el camino que nos queda por recorrer es maravilloso. También quiero agradecer las amistades tan fuertes que he forjado estos últimos años, personas que me han acompañado y a las que aprecio profundamente. De igual forma querría agradecer a Agapito Ledezma su apoyo y trabajo, gracias al cual he podido finalizar este proyecto.

Por último, quería agradecer a la universidad Carlos III todos los momentos felices y enseñanzas que me ha dado, considero que las asociaciones de la universidad son un bien a proteger que forman a los alumnos en el compañerismo, amistad y el trabajo en equipo, valores necesarios en el desarrollo de las personas que no se consiguen yendo solo a clase.

Glosario de términos

csv	<i>Comma Separated Values</i>
ECH	Encuesta Continua de Hogares
IDE	Entorno de Desarrollo Integrado
EPA	Encuesta de Población Activa
INE	Instituto Nacional de Estadística
MaaS	Movilidad como un servicio
MAS	Sistemas Multi-Agente
Matsim	<i>Multi-Agent Transport SIMulation</i>
MBA	Modelado Basado en Agente
OSM	OpenStreetMap
TRANSIMS	<i>TRansportation ANalysis SIMulation System</i>
WGS84	World Geodetic System 84
XML	<i>eXtensible Markup Language</i>

Índice de contenidos

1. Introducción	15
1.1. Motivación	15
1.2. Objetivos	15
1.3. Marco regulador.....	15
1.4. Estructura del Documento	16
2. Estado del arte	18
2.1. Congestión de las ciudades	18
2.1.1. Aumento de la población; Proyección de futuro	18
2.1.2. Desruralización.....	20
2.1.3. Ciudades del futuro	21
2.2. Evolución de los transportes	23
2.2.1. Los transportes en la actualidad	24
2.2.2. Los transportes en un futuro próximo	24
2.3. El coche autónomo.....	25
2.4. Agentes.....	27
2.4.1. Arquitecturas de agente.....	28
2.4.2. Propiedades de los agentes	29
2.4.3. Nuevas tecnologías a partir de los agentes.....	30
2.5. Sistemas multiagente (MAS)	30
2.5.1. Características	30
2.5.2. Ventajas de los MAS.....	31
2.5.3. Propiedades.....	31
2.5.3.1. Competencias.....	31
2.5.3.2. Modelos de agentes conocidos.....	32
2.5.3.3. Comunicación entre agentes.....	32
2.5.3.4. Comportamiento.....	32
2.5.3.5. Puntos de interacción	32
2.5.4. Aplicaciones.....	33
2.6. Simulación basada en agentes	33
2.7. MATSIM.....	34
2.7.1. Historia y Bases	34
2.7.2. Algoritmo Coevolutivo	35
2.7.3. Esquema de funcionamiento	36

2.7.4. Modelo de Flujo del Trafico	36
3. Diseño de la solución.....	38
3.1. Generación de la red	38
3.1.1. OpenStreetMap.....	39
3.1.2. JOSM.....	39
3.2. Generación de la población.....	42
3.2.1. Sintetizadores de población.....	43
3.2.2. Nuevo enfoque en la sintetización de población; Estructuras sociales	46
3.2.3. Descripción del programa: <i>The Synthetic Population</i>	47
3.2.4. Datos y Fuentes	49
3.2.4.1. Censo de Leganés	49
3.2.4.2. Datos estadísticos poblacionales del INE	49
3.2.4.3. Datos de los inmuebles del Catastro.....	50
3.2.4.4. Datos de los desplazamientos de la encuesta de movilidad del CRTM	51
3.2.5. Creación de las familias: El Familiador	52
3.2.5.1. Estudio sociológico de los hogares.....	53
3.2.5.2. Tipologías de hogares.....	56
3.2.5.3. Estructura y funcionamiento.....	60
3.2.5.4. Selección del género y edad: características familiares.....	64
3.2.5.5. Mejoras	67
3.2.6 Generación de los planes: El Planificador	69
3.2.6.1. Clasificación de los hogares reales	70
3.2.6.2. Estructura y funcionamiento.....	72
3.2.6.3. Ráster de Leganés	74
3.2.6.4. Preprocesado de la población real.....	80
3.2.6.5. Traspaso de planes y escritura XML.....	83
3.2.6.6. Obtención de nuevas coordenadas.....	84
3.2.6.7. Mejoras	91
4. Análisis y evaluación del sistema	93
4.1. Casos de uso	93
4.1.1. Tabla de requisitos y nomenclatura.....	93
4.1.2. Especificación de Casos de Uso.....	93
4.2 Requisitos de software	94
4.2.1. Tabla de requisitos y nomenclatura	94
4.2.2 Especificación de Requisitos funcionales	95
4.2.3. Especificación de Requisitos no funcionales	101

4.3 Evaluación	103
5. Simulación	105
5.1 Simulación del proyecto	107
6. Gestión del proyecto	110
6.1 Planificación	110
6.2 Presupuesto	111
6.2.1 Costes del personal	111
6.2.2 Costes del equipo	113
6.2.3 Costes para fungibles y otros gastos varios	114
6.2.4 Costes totales del proyecto	114
6.3 Impacto socio-económico	114
7. Conclusiones y trabajos futuros	116
7.1 Conclusiones.....	116
7.2 Trabajos futuros	117
Bibliografía	118
Anexo A	122
A.1. Datos y Fuentes.....	122
A.1.1. Censo de Leganés.....	122
A.1.2. Datos estadísticos poblacionales del INE.....	123
A.1.3. Datos de los inmuebles del Catastro	126
A.1.4. Datos de los desplazamientos de la encuesta de movilidad del CRTM	127
Anexo B	130
B.1. Utilización del software.....	130
Anexo C	132
C.1. Introducción	132
C.2. State of art	133
C.2.1. City congestion.....	133
C.2.2. Evolution of transport	134
C.2.3. Autonomous car.....	134
C.2.4. Agents.....	135
C.2.5. Multi-agent systems.....	135
C.2.6. Agent-based model	136
C.2.7. MATSIM.....	136
C.3. Diseño de la solución	137
C.3.1. Network generation	137
C.3.2. Generating the population.....	138

C.3.2.1. The Synthetic Population	139
C.3.2.2. Generation of families, El Familiador.....	139
C.3.2.3. Generation of plans. El Planificador.....	140
C.4. Evaluation.....	142
C.5. Conclusions	144

Índice de figuras

Figura 1. Evolución de la población mundial	18
Figura 2. AFP - Previsiones demográficas en África	20
Figura 3. Densidad de población en Indonesia	22
Figura 4. Reconstrucción de la ciudad de Çatalhöyük	23
Figura 5. Ejemplo arquitectura subsunción	28
Figura 6. Interacción de los agentes	29
Figura 7. Visualización de la simulación en Matsim	35
Figura 8. Bucle funcionamiento de Matsim	36
Figura 9. Modelo de flujo del tráfico	37
Figura 10. Mapa de los procesos a implementar para poder simular	38
Figura 11. Mapa de procesos a implementar para obtener la red	39
Figura 12. Mapa de Leganés descargado de OpenStreetMap	40
Figura 13. Mapa de la red viaria de Leganés obtenida con el <i>plugin</i>	41
Figura 14. Tabla del <i>plugin</i> con datos de las vías	42
Figura 15. Diagrama The Synthetic Population.....	48
Figura 16. Tabla de los usos de los bienes inmuebles	50
Figura 17. Tabla hogares por tipo y número de personas	57
Figura 18. Diagrama tipologías de hogar 1	58
Figura 19. Diagrama tipologías de hogar 2	59
Figura 20. Diagrama del Familiador	61
Figura 21. Diagrama del Planificador	73
Figura 22. Ráster de Leganés	74
Figura 23. Conversión tipos de actividad	76
Figura 24. Celda del ráster	76
Figura 25. Diagrama submódulo 1	79
Figura 26. Diagrama organización de la población real.....	81
Figura 27. Diagrama submódulo 2	82
Figura 28. Diagrama submódulo 3	85
Figura 29. Mapa de desplazamientos INE	86
Figura 30. Puntos de salida del municipio de Leganés	87
Figura 31. Cálculo del nuevo destino 1	87
Figura 32. Cálculo del nuevo destino 2	88
Figura 33. Cálculos del nuevo destino 3	88
Figura 34. Generación de nuevas coordenadas 1	89

Figura 35. Generación de nuevas coordenadas 2	89
Figura 36. Generación de nuevas coordenadas 3	89
Figura 37. Diagrama submódulo 4	90
Figura 38. Comparativa entre tipologías de hogar reales y sintéticas	103
Figura 39. Actividades de la población sintética	104
Figura 40. Actividades de la población real	104
Figura 41. Matsim en Eclipse	106
Figura 42. GUI Matsim	106
Figura 43. Histograma en la iteración 7	107
Figura 44. Histograma en la última iteración	108
Figura 45. Puntuaciones por iteración	108
Figura 46. Simulación de Leganés en Matsim	109
Figura 47. Diagrama de Gantt	112
Figura 48. Selección de variables en el generador de tablas a medida	124
Figura 49. Selección de filtros en el generador de tablas a medida	125
Figura 50. Tabla de los usos de los bienes inmuebles	127
Figure 51. Interaction of agents	135
Figure 52. Loop operation of Matsim	136
Figure 53. Map of the processes to be implemented in order to be able to simulate	137
Figure 54. Map of the road Leganés network of obtained with the plugin	138
Figure 55. Calculation of the new destination	142
Figure 56. Comparison between real and synthetic home typologies	143
Figure 57. Synthetic population activities	143
Figure 58. Real population activities	143

Índice de tablas

Tabla 1. Cambio de orígenes.....	51
Tabla 2. Cambio municipio.....	52
Tabla 3. Porcentajes de hogar por número de personas.....	57
Tabla 4. Agrupación de estadísticas por rangos	62
Tabla 5. Diferencia de edad en la pareja	65
Tabla 6. Plantilla tabla casos de uso	93
Tabla 7. Caso de uso CU-01.....	93
Tabla 8. Caso de uso CU-02.....	94
Tabla 9. Caso de uso CU-03.....	94
Tabla 10. Plantilla tabla requisitos de software.....	94
Tabla 11. Requisito de software RS-RF-01	95
Tabla 12. Requisito de software RS-RF-02	96
Tabla 13. Requisito de software RS-RF-03	96
Tabla 14. Requisito de software RS-RF-04	96
Tabla 15. Requisito de software RS-RF-05	97
Tabla 16. Requisito de software RS-RF-06	97
Tabla 17. Requisito de software RS-RF-07	97
Tabla 18. Requisito de software RS-RF-08	98
Tabla 19. Requisito de software RS-RF-09	98
Tabla 20. Requisito de software RS-RF-10	98
Tabla 21. Requisito de software RS-RF-11	99
Tabla 22. Requisito de software RS-RF-12	99
Tabla 23. Requisito de software RS-RF-13	99
Tabla 24. Requisito de software RS-RF-14	100
Tabla 25. Requisito de software RS-RF-15	100
Tabla 26. Requisito de software RS-RF-16	100
Tabla 27. Requisito de software RS-RF-17	101
Tabla 28. Requisito de software RS-RF-18	101
Tabla 29. Requisito de software RS-RF-19	101
Tabla 30. Requisito de software RS-NF-01.....	102
Tabla 31. Requisito de software RS-NF-02.....	102
Tabla 32. Requisito de software RS-NF-03.....	102
Tabla 33. Requisito de software RS-NF-04.....	102
Tabla 34. Requisito de software RS-NF-05.....	103

Tabla 35. Planificación	111
Tabla 36. Costes de personal	113
Tabla 37. Costes del equipo	113
Tabla 38. Costes fungibles y otros	114
Tabla 39. Costes totales del proyecto	114
Tabla 40. Cambio de orígenes.....	128
Tabla 41. Cambio municipio.....	129
Tabla 42. XML censo	131
Tabla 43. XML catastro	131
Tabla 44. XML CRTM	131

1. Introducción

1.1. Motivación

Los grandes cambios que se están produciendo en esta era tecnológica, más palpables que nunca tras la crisis del coronavirus, han acelerado la digitalización de todos los campos y están modificando la forma en que las personas interactúan y se mueven. Las ciudades que acogen a esta población cada vez mayor, van a tener que reinventarse para dar respuesta a todas estas necesidades.

La mejor forma de estudiar estos cambios y encontrar soluciones viene de la mano de la simulación, este trabajo pretende facilitar el acceso a ésta, sirviendo de base para futuros trabajos y estudios que ayuden al desarrollo de soluciones.

1.2. Objetivos

El objetivo principal de este trabajo busca sentar las bases para simulaciones de transporte urbano complejas, preparando y configurando un entorno de simulación basado en agentes de modo que se garantice una simulación representativa de la realidad, facilitando el trabajo a futuros usuarios.

Para cumplir este objetivo se describirán y desarrollarán todos los procesos y los tipos de datos necesarios, con los que construir una base sólida sobre la que se pueda simular. Los objetivos específicos orientados a conseguir el principal son los siguientes:

1. Recopilar los tipos de datos necesarios.
2. Citar posibles fuentes de las que se puedan obtener datos relevantes.
3. Generar una red viaria sobre la que poder simular.
4. Generar una población sintética.
5. Generar planificaciones diarias y asignárselas a las personas sintéticas.

Para generar la población sintética, este TFG propone un método basado en variables sociológicas extraídas de la población real, utilizando las estructuras sociales básicas que definen el comportamiento y la composición de los hogares.

1.3. Marco regulador

En este apartado se presentan los aspectos legales y técnicos que se deben tener en cuenta a la hora de implementar el sistema desarrollado.

Dado que este proyecto tiene una aplicación a nivel teórico, no existe una regulación al respecto en el código legal español. Sin embargo, para la creación del programa se han utilizado datos de personas reales cuya protección de datos, si es contemplada en el código legal español en la ley orgánica 3/2018, del 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales.

Los datos recogidos provienen de tres fuentes principales, siendo las tres públicas, el INE, el ayuntamiento de Leganés y el Consorcio Regional de Transportes de Madrid. La base jurídica para el uso de datos por parte de estas entidades viene recogida en el artículo 6.1.e) del RGPD:

“Tratamiento necesario para el cumplimiento de una misión realizada en interés público o en el ejercicio de poderes públicos conferidos al responsable del tratamiento”.

Los datos recogidos de estas fuentes no contienen la información necesaria con la que poder identificar a los individuos. Estos datos son convenientemente anonimizados para preservar la confidencialidad. Por ello, este proyecto no incurre en problemas de protección de datos.

El software utilizado para este trabajo es de código abierto, aun así, estos programas cuentan con diferentes licencias para regular su uso, copia, modificación y/o distribución. Las licencias de los programas utilizados en este trabajo son las siguientes:

- **Matsim:** Este software está licenciado bajo los términos de *GNU General Public License v2.0*, permitiendo la copia y la distribución del programa, pero no la modificación del mismo. Para más información sobre los términos de la licencia se remite al documento de *GNU* [1].
- **OpenStreetMap:** Este software cuenta con la licencia *Open Data Commons Open Database License (ODbL)* de la Fundación *OpenStreetMap* (OSMF) y permite a los usuarios copiar, distribuir, transmitir y adaptar los datos siempre que se de reconocimiento a *OpenStreetMap*. Para más información sobre los términos de esta licencia se remite al documento de *Open Data Commons* [2].

1.4. Estructura del Documento

Para favorecer la comprensión del documento este punto contiene su estructura, a su vez se presentarán brevemente los contenidos de cada apartado.

- **Capítulo 1:** Introducción al proyecto que abarca los objetivos principales, las motivaciones y los aspectos legales.
- **Capítulo 2:** Estado del arte. Presenta la situación actual sobre la que parte el trabajo, desarrollando aspectos relativos a la población, a las ciudades, a la modelización basada en agentes y al simulador escogido, Matsim.
- **Capítulo 3:** Diseño de la solución, se incluyen las necesidades previas a la simulación, desarrollando su solución mediante el diseño de un software.
- **Capítulo 4:** Análisis del sistema, contiene la especificación de los casos de uso, de los requisitos del sistema y un análisis de los resultados obtenidos del programa desarrollado.
- **Capítulo 5:** Experimentación, presenta los procesos básicos para trabajar con Matsim a la vez que muestra el correcto funcionamiento del programa con los archivos obtenidos en el capítulo 3.
- **Capítulo 6:** Gestión del proyecto, recoge la planificación seguida durante el proyecto e indica el presupuesto previsto y el impacto socioeconómico esperable gracias a la creación del software de este trabajo.

- **Capítulo 7:** Conclusiones y desarrollo futuro, se incluyen todas las reflexiones finales obtenidas tras la finalización de este trabajo fin de grado. Incluye las líneas de desarrollo futuras que se derivan de éste.
- **Anexo A:** Extensión del apartado Datos y fuentes, donde se desarrollan pormenorizadamente los procesos de extracción y preprocesado realizados a los datos.
- **Anexo B:** Guía de usuario de The Synthetic Population, que recoge los datos estadísticos necesarios para el uso del programa, así como las estructuras de los archivos de datos que se tienen que pasar, para el correcto funcionamiento del mismo.
- **Anexo C:** Competencias obligatorias en inglés, se incluye un resumen del trabajo en su totalidad.

2. Estado del arte

Este punto muestra la coyuntura actual sobre la que arranca este proyecto, estudiando la situación de las ciudades. Para ello, se ha analizado el crecimiento de las urbes debido al aumento de la población y a la desruralización, para entender como está cambiando la movilidad de las personas. Adicionalmente se han estudiado los nuevos medios de transportes, haciendo especial hincapié en los coches autónomos, y los nuevos sistemas que gestionan los desplazamientos que se engloban en la *Smart Mobility*. Por otra parte, se han expuesto los conceptos agente, sistema multiagente y modelado basado en agentes, explicando su composición y funcionamiento. Por último se ha presentado Matsim, el simulador escogido para este proyecto.

2.1. Congestión de las ciudades

Para entender por qué es necesaria la simulación de la movilidad tanto en la actualidad como para un futuro próximo, este apartado va a enumerar las distintas causas que están cambiando la situación actual de las ciudades, señalando como la simulación va a ser clave para entender estos cambios y para saberse adaptar a ellos.

2.1.1. Aumento de la población; Proyección de futuro

Las revoluciones industriales introdujeron mejoras en muchos campos, que trajeron consigo grandes avances, como el aumento de la producción de alimentos, o las mejoras en el campo de la medicina. Esto hizo posible mejorar las condiciones de vida de las personas, produciéndose así una explosión poblacional, que empezó en el siglo XX y que gracias a los continuos avances que se han ido dando, ha continuado hasta la actualidad.

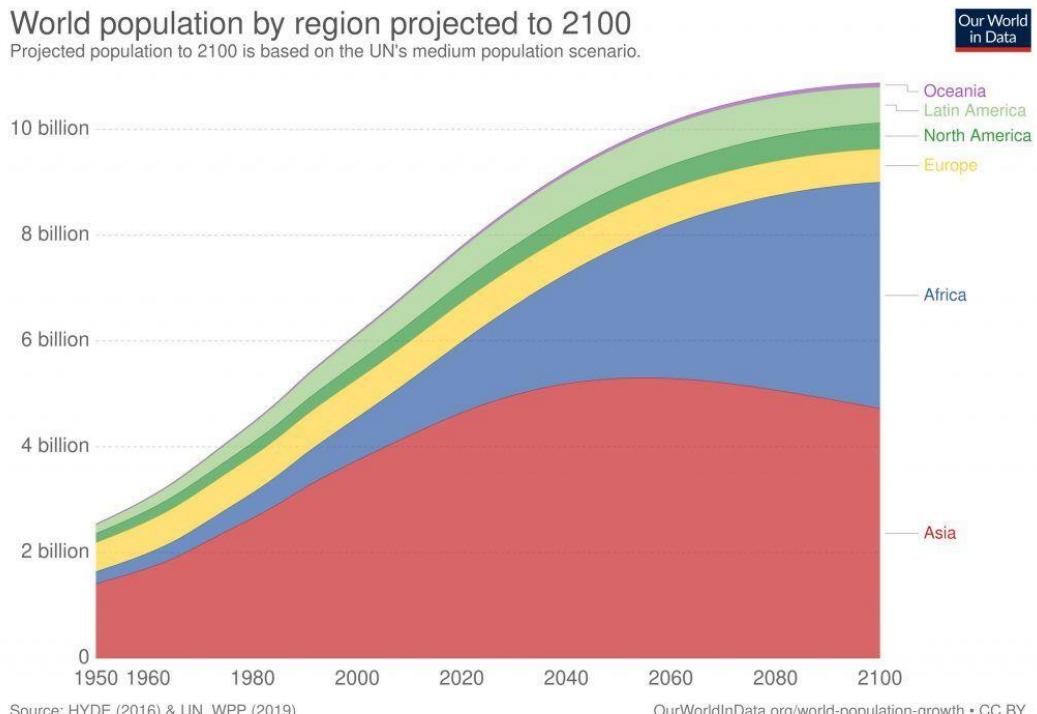


Figura 1. Evolución de la población mundial [6]

La población mundial, que ha ido aumentado exponencialmente, está viendo las primeras señales de estancamiento en los países desarrollados [3], con Europa, Oceanía y América del norte a la cabeza. La natalidad ha disminuido drásticamente [4] y esto es debido múltiples factores, como el aumento del precio de la vida que hace que algunas familias no puedan asumir los gastos que supone tener un hijo. Según la OCU el gasto medio durante el primer año de vida de los hijos es de 7.706€¹ [5]. Por otro lado, la educación de la mujer y su inserción al mundo laboral ha complicado la conciliación entre la maternidad y el desarrollo de su carrera profesional.

Todo esto junto a los métodos anticonceptivos ha desembocado en una reducción drástica de la natalidad y, por tanto, de un envejecimiento de la población en el continente europeo, donde solo la inmigración hace frente a esta bajada demográfica.

En el caso de Asia, que actualmente representa el 61% de la población mundial, se espera que, aunque sus números sigan aumentando a gran velocidad, en torno al año 2060 este crecimiento se estanke y solo llegue a representar el 54% de la población humana². Esto es debido a que la fecundidad ha caído en la mayoría de los países asiáticos. Para que una generación reemplace a otra, las mujeres deben tener al menos 2 hijos, esto queda representado por el índice de fecundidad que debe ser mayor a 2.1 para que se produzca el reemplazo generacional. Sin embargo, esto no se está cumpliendo, pues muchos países asiáticos están rozando el límite necesario para el reemplazo, mientras que, en otros muchos, este índice ya se encuentra por debajo del nivel de reemplazo. Esta bajada de la natalidad no va a suponer un descenso directo de la población debido a la inercia demográfica, esto significa que, si hay mayor población joven en edad de tener hijos que población muy envejecida, aunque los jóvenes tengan menos hijos que antes, las defunciones se darán a un ritmo más bajo que los nacimientos. Esto es lo que hace que la población asiática siga creciendo en gran medida, pero con una “fecha de caducidad” para mitad de siglo.

Pese a que la situación de Asia en conjunto es de crecimiento hasta 2060, la diferencia entre unos países del continente asiático y otros es abismal. Países como la India, Pakistán o Indonesia están en la cabeza del crecimiento poblacional, mientras que en el otro extremo están China, Corea del Sur y Japón, donde la natalidad es muy baja y la inercia demográfica no se produce. China ha intentado parar esta situación con la derogación de la política del hijo único en 2015, sin obtener el resultado esperado, pues, aunque las parejas chinas ya pueden tener dos hijos, la incorporación de la mujer al trabajo y los altos costes de la vida hacen que pocas familias opten por concebir un segundo hijo [7].

En el lado demográfico opuesto está el continente africano [8], con un índice de natalidad muy superior al resto del mundo y aunque esta cifra se ha reducido los últimos años, las mujeres siguen teniendo 4.7 hijos de media. Este hecho es debido a múltiples factores, como el escaso uso de métodos anticonceptivos, (tanto por la falta de educación de la población, como por la falta de acceso a ellos), o la elevada cifra de embarazos en adolescentes, (al tener los hijos a tan temprana edad, el periodo de fertilidad es más amplio, pudiéndose dar un mayor número de embarazos a lo largo de su vida). Otra razón que hace que el índice de natalidad sea tan alto, es que, culturalmente, las familias africanas valoran positivamente tener un mayor número hijos. Todo esto hace que, para 2050, se espere que la población africana se duplique [9].

¹ Segundo un Estudio realizado por la OCU

² Porcentaje calculado a partir de datos obtenidos de populationpyramid.net

Africa's demographic explosion

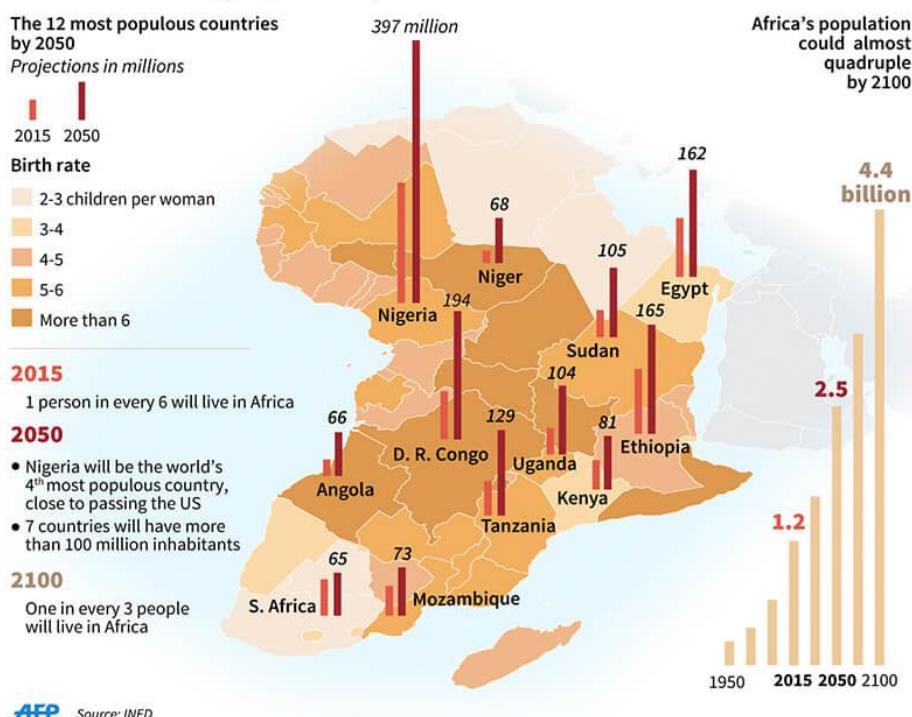


Figura 2. AFP - Previsiones demográficas en África [8]

Dentro del continente africano hay grandes diferencias. Mientras que los países árabes al norte del Sahara y los países más meridionales tienen menores previsiones de crecimiento, los países del África occidental y sobre todo del centro de África tienen unas expectativas de crecimiento muy elevadas. Nigeria es el país que se sitúa a la cabeza de esta explosión poblacional, pues se espera que el actual país más poblado de África se convierta en el tercer estado más poblado del mundo en 2050.

Vista la situación en los distintos continentes y pese a estas señales de estancamiento en el crecimiento poblacional del planeta, la población mundial está a punto de llegar a los 8 mil millones de personas, y se acerca cada vez más a la cifra redonda de los 10 mil millones, una cifra peligrosa que podría causar una escasez de alimentos y de otros tantos recursos, un planeta que podría colapsar ante una población mundial tan elevada. No obstante, algunos estudios como el publicado por The Lancet [10] indican que la población no va a alcanzar la temida cifra de los 10 mil millones, pues se obtendrá un pico en la década de 2060 con 9.700 millones de personas. Punto tras el cual, debido a la educación de la mujer y el mayor acceso a los métodos anticonceptivos, derivaría en un decrecimiento de la población mundial.

Aunque no es seguro que se vayan a alcanzar los 10 mil millones de personas, lo que si queda claro es que la población actual va a seguir creciendo a gran ritmo y con éste.

2.1.2. Desruralización

Además del fenómeno del aumento de la población comentado el apartado anterior, se está produciendo una desruralización. Este es un fenómeno que se está produciendo a nivel mundial, dado que la tasa de urbanización de un país está directamente relacionada con el desarrollo económico de éste. En la medida en que los países en vías de desarrollo crecen económicamente, su población rural se va concentrando en las ciudades. A grandes rasgos, esto

es debido al cambio en el tejido económico, con una bajada del peso del sector agrícola en la economía del país y con la diversificación económica. Las ciudades se industrializan, a la vez que aumentan sus servicios para acoger a esta nueva población inmigrante, lo que hace que se dé una mayor productividad en las áreas urbanas.

Cada vez más personas abandonan el campo hacia los núcleos urbanos en busca de un futuro mejor. Siendo alrededor de 2014 [11] cuando por primera vez la población que vivía en las ciudades superó a la población rural. Pero aparte de la relación ya comentada entre el desarrollo económico y la urbanización de un estado, ¿qué es lo que hace que las personas individualmente tomen la decisión para dejar el campo? Son muchos los factores que están produciendo esta migración hacia las urbes, como la mecanización del campo, que reduce la necesidad de mano de obra poco cualificada, la diferencia salarial entre las ciudades y el campo, la cantidad y calidad de los servicios presentes en las metrópolis, la desaparición en los pueblos de la gente joven que es más propensa a emigrar y que, por tanto, genera un déficit de parejas para los que se han quedado. Toda esta población emigrante genera un efecto llamada, cuando a un migrante le va bien en la ciudad, otras personas de su antiguo entorno rural, que en un principio no se atrevieron a dar el paso, deciden finalmente darlo. De esta manera, se produce un círculo vicioso que se va retroalimentando, la disminución de la población de los núcleos rurales hace que este entorno se vuelva menos atractivo, lo que a su vez hace que más gente abandone estos pueblos. Esto está llevando a la desaparición de pueblos que ven como los jóvenes se van, quedándose en ellos solo la gente más mayor, que va poco a poco muriendo y con ellos el pueblo.

Frente a este fenómeno indeseable, aún queda alguna pequeña esperanza para la supervivencia de las zonas rurales, que pasa por la transformación de éstas, en zonas “miniurbanas” que sepan reinventarse. Para que un pueblo sobreviva, éste debe ofrecer algún incentivo que atraiga a nuevas personas a mudarse a él. Esto pasa por la correcta gestión de los ingresos obtenidos a causa de la modernización del sector agrario, que ha conseguido mayores excedentes. Estos nuevos ingresos deberían ser reinvertidos en el pueblo, generando con ellos más y nuevos servicios que atraigan a nueva población, concentrando así los servicios de la comarca y convirtiéndose por ello en el pueblo de referencia de la zona. A esto se le suma que, con la digitalización, el aumento del teletrabajo y la entrada del 5G, se hace posible que personas del entorno urbano puedan mudarse al campo y teletrabajar desde éste. Otro factor que juega a favor de los entornos rurales viene de la mano de la burbuja inmobiliaria de las ciudades, que hace imposible que los jóvenes puedan permitirse una vivienda en ellas. Todos estos factores dan la esperanza de revertir en un futuro este proceso de abandono de los pueblos. Sin embargo, en la situación actual este proceso de abandono del campo continua.

2.1.3. Ciudades del futuro

Para el año 2050 se estima que el 68% de la población vivirá en las ciudades [12]. Con la concentración de la sobre población en las metrópolis, la congestión de las ciudades, que no fueron diseñadas para soportar tantas personas, es inevitable. Tal como se ha visto en la sección de población, se van a masificar ciudades no planificadas, ciudades sin recursos ni infraestructuras, especialmente en el continente africano donde se van a multiplicar en el futuro sus ya numerosas deficiencias.

En estas circunstancias será de vital importancia, que se estudien los problemas, para ver cómo se pueden revertir todas estas carencias. Un ejemplo claro y actual de lo que puede llegar a suceder en estas futuras urbes se puede observar en Yakarta. Esta ciudad dejará de ser la capital de Indonesia tras el traslado a una nueva capital construida desde cero, este cambio de capital es debido al claro fracaso de las políticas, que no han sabido dar respuesta a los problemas de la ciudad.

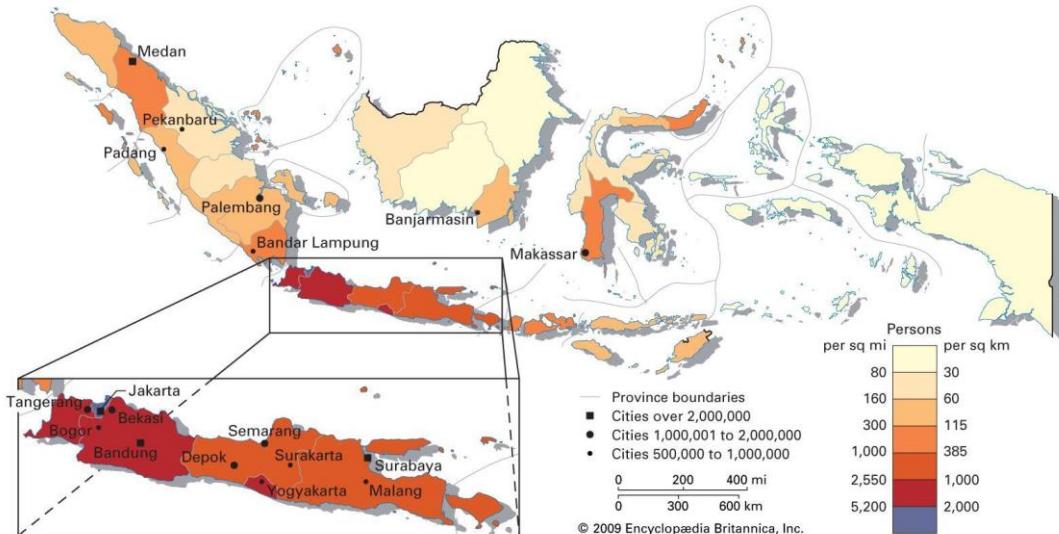


Figura 3. Densidad de población en Indonesia [13]

Yakarta ha sido y sigue siendo un centro captador de población que concentra la mayor parte de la economía y de la población del país. Una ciudad que fue creciendo sin control, donde las personas que llegaban creaban sus propias casas, generándose extensas barriadas que no tenían acceso a los servicios básicos, como alcantarillado o agua corriente. Esto ha provocado que las familias hayan creado pozos para obtener agua de forma ilegal, lo que ha producido que la ciudad se hunda 15 centímetros al año. Por otro lado, la situación del tráfico se ha ido agravando con el crecimiento de la ciudad, la falta de planificación y las malas infraestructuras hacen muy difícil el desplazamiento. Todo esto junto a otras razones, como la necesidad de desarrollar otras partes del país, ha desembocado en la creación de una nueva capital en la isla de Borneo.

Por otro lado, en los países desarrollados, la situación es diferente, no se prevé un aumento significativo de la población en su conjunto e incluso, en algunos casos, podrían perderse habitantes. Sin embargo, en algunas grandes ciudades, como consecuencia de la continua inmigración desde las zonas rurales o desde terceros países, la población seguirá aumentando, como en el caso de Madrid o de Roma. De cualquier manera, en todas las ciudades, aumenten o disminuyan su población, tienen otro problema común, el envejecimiento de su población. Este hecho va a provocar nuevas necesidades en estas futuras ciudades, dado que, una población envejecida va a propiciar que se den ciertos cambios en los hábitos vitales de los ciudadanos, lo que precisará de una adaptación de las urbes a las nuevas circunstancias.

Para buscar soluciones a los problemas de las ciudades actuales, muchas grandes metrópolis se están digitalizando, surgiendo con ello el concepto de las *Smart Cities* [14]. Este concepto abarca a todo lo que rodea a las ciudades, empleando las tecnologías de la información y de la comunicación (TIC) con el objetivo de mejorar la calidad de vida de sus habitantes. Garantizando con ello un desarrollo sostenible, que optimiza los recursos disponibles y facilita la participación de la ciudadanía activa. Para ello se hace uso del Big Data, recogiendo distintos tipos de datos de los ciudadanos, de las vías, del medio ambiente, etc. mediante sensores. Para posteriormente procesarlos, pudiendo con ello identificar problemas.

La aplicación del concepto de *Smart Cities* pretende mejorar las ciudades actuales en campos como la economía, el gobierno, el medio ambiente, la cohesión social o la movilidad dentro de ellas. Es en este último punto, donde este proyecto pretende aportar medios e ideas con los que apoyar el desarrollo de las *Smart Cities* [15].

Cabe comentar que la introducción de las mejoras planteadas por las ciudades inteligentes y en concreto las que estudian la movilidad, podrá ser aplicado con mayor facilidad en los países desarrollados, donde la inversión en estas materias es mucho más amplia y donde la introducción del coche autónomo ya es una realidad. Mientras que, en otros países con gobiernos menos organizados o con pocos medios para invertir en infraestructuras, la tarea no va a ser nada fácil. De cualquier manera, el futuro para todos pasa por detectar problemas y estudiar sus soluciones, utilizando para ello simulaciones que estudien entre otros la circulación.

2.2. Evolución de los transportes

Inicialmente, las ciudades estaban pensadas para las personas, el objetivo de éstas era el de darles un hogar a sus habitantes. Dado que la mayoría de los transeúntes eran personas, las calles eran estrechas, llegando incluso a crearse ciudades como Çatalhöyük³, en las que no existían calles y el acceso a las casas se daba por los techos. El aumento del comercio creó nuevas necesidades, haciendo que fuera indispensable la construcción de grandes vías para que los carros de los comerciantes tirados por caballos o mulas pudieran pasar. Por otro lado, también la necesidad de mover y organizar los ejércitos dentro de la ciudad, propicio que se crearan estas amplias calles. Muchos siglos después las ciudades tuvieron que adaptar sus redes viarias para la introducción del automóvil y, aunque al principio los coches eran pocos y las personas eran mayoría, con el tiempo, el aumento del tráfico hizo que los vehículos tomaran las calles. Un proceso que se ha ido revirtiendo poco a poco una vez entrados en el siglo XXI, con la peatonalización de los cascos históricos.

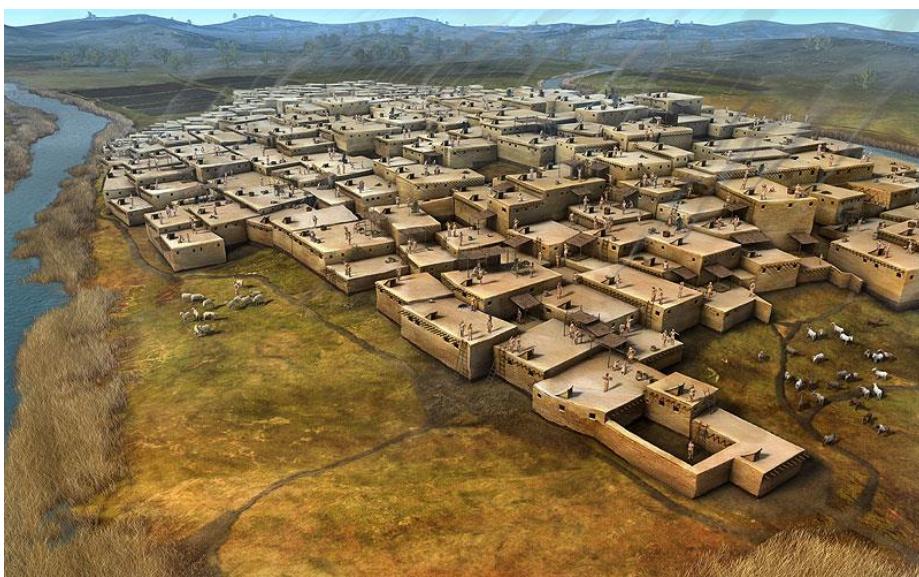


Figura 4. Reconstrucción de la ciudad de Çatalhöyük [16]

Como ha quedado plasmado, las ciudades siempre han estado en continuo cambio, sabiéndose adaptar a las nuevas necesidades que iban surgiendo. En la actualidad la hiperpoblación, la globalización y los cambios en las formas de consumo están creando nuevas necesidades, para poder satisfacer estas necesidades han surgido nuevas formas de desplazarse como los patinetes eléctricos, el *car-sharing*, el coche autónomo o nuevas formas de gestionar los desplazamientos en conjunto como la movilidad inteligente, dando soluciones a los problemas

³ Çatalhöyük fue un asentamiento poblado durante el neolítico en torno al 6000 a. C. que se sitúa en la actual Turquía

de las urbes. No obstante, la introducción de éstas implica, como tantas otras veces a lo largo de la historia, que las ciudades tengan que adaptarse a estas nuevas tecnologías.

2.2.1. Los transportes en la actualidad

El crecimiento demográfico de las ciudades ha provocado una curva ascendente en el número de desplazamientos, lo que supone una saturación de las infraestructuras de las urbes. Por otro lado, la globalización ha unido al mundo, generando una cadena de transportes que conecta las materias primas con las fábricas y con el consumidor final. Una cadena muy compleja que afecta especialmente a la movilidad de las ciudades en su tramo final y es que el consumidor ha cambiado drásticamente sus hábitos, donde los desplazamientos individuales para comprar se han ido reduciendo, primando al comercio electrónico que con la distribución domiciliaria que facilita el día a día de los usuarios, otros servicios como los de *food delivery* siguen esta estela de la comodidad del hogar y están en plena expansión. Aunque este modelo de entrega a domicilio ha supuesto un descenso del tráfico privado, a su vez ha aumentado los desplazamientos para el reparto, lo que ha supuesto un reto para estas empresas que buscando reducir los costos y los tiempos de entrega, han tenido que invertir gran parte de sus recursos para conseguir la máxima eficiencia en el reparto. Gracias a todos estos estudios realizados, se han adquirido unos conocimientos de reorganización estratégica, táctica e informática del movimiento de mercancías que progresivamente se están aplicando a la movilidad de pasajeros.

En todos los transportes, pero especialmente en el interior de las ciudades, han surgido nuevos tipos de transporte que han solapado sus servicios con otros más tradicionales. Ofreciendo nuevos servicios como el *car-sharing* y los patinetes eléctricos, o dándoles un lavado de cara a servicios ya existentes como los VTC, todos comparten la inmediatez, que funcionan mediante *apps* y que se oponen al transporte público colectivo, generando un mayor número de viajes privados.

Hasta ahora, ante el crecimiento de las ciudades, la solución a la congestión del tráfico pasaba por construir más carreteras o por aumentar la capacidad de las vías preexistentes. Sin embargo, esto ha conducido a un aumento del tráfico, pues con la ampliación de la capacidad, se incentiva y orienta a la población a hacer uso del coche privado. Medidas como la creación de zonas de aparcamiento disuasorio, no han conseguido reducir la cantidad de movilidad privada objetivo. Otra medida que se está introduciendo para fomentar el uso del transporte público es la peatonalización del centro de las ciudades, pero ésta no está mostrando todo su potencial por culpa de la pandemia, que hace que mucha gente no quiera coger los trenes o los buses por miedo al contagio. Buscando alcanzar lo que no han podido conseguir las medidas mencionadas, surge la *Smart Mobility*, la cual busca una solución inteligente a los problemas de movilidad, sin centrarse en aumentar las capacidades de las vías o en el establecimiento de nuevas zonas de aparcamiento disuasorio.

2.2.2. Los transportes en un futuro próximo

La *Smart Mobility* o movilidad inteligente [15] es uno de los elementos de los que se componen las *Smart Cities*. Esta rama estudia cómo conseguir una movilidad más limpia, segura y eficiente, reduciendo la contaminación y las pérdidas de tiempo asociadas al tráfico. Se fomenta con ello el uso del transporte público y de los vehículos autónomos. La movilidad inteligente abarca todos los medios de transporte conocidos actualmente, con la diferencia de que estos dejan de ser propiedad del usuario. De esta manera, se aleja la idea del vehículo como transporte personal propio, e introduce la idea de que este vehículo es un servicio del que disponen las personas. A esta forma de movilidad se la conoce por movilidad como un servicio (MaaS)

Un trayecto se compondrá de una combinación de transportes, tanto privados como públicos, que serán gestionados y unificados mediante una puerta de enlace para conseguir realizar el trayecto puerta a puerta. Esta movilidad inteligente se rige por los siguientes principios [17]:

- **Flexibilidad:** Al existir múltiples medios de transporte, es el usuario el que escoge cual es el que mejor se ajusta a sus necesidades.
- **Eficiencia:** El recorrido ofrecido debe producirse en el menor tiempo posible, evitando interrupciones durante el trayecto.
- **Integración:** El trayecto se planifica aunando los distintos medios de transporte que se vayan a utilizar.
- **Tecnología limpia:** Se busca que los distintos modos de transportes utilizados sean lo más ecológicos posibles, tendiendo hacia las cero emisiones.
- **Seguridad:** Con las nuevas tecnologías y el uso del coche autónomo, se busca reducir significativamente las muertes y accidentes en los viajes, siendo el coche autónomo o la movilidad inteligente soluciones a este problema.

La propuesta de los MaaS pretende dar solución al conjunto de problemas planteados en las ciudades actuales, consiguiendo una serie de ventajas [18].

Para lograr esta mejora es imprescindible avanzar en el grado de autonomía de los vehículos y desarrollar sistemas MaaS, de modo que tanto el uso de los vehículos como la implementación de las decisiones y planes de desplazamiento se hagan del modo más coordinado y rentable posible.

2.3. El coche autónomo

Con la introducción del coche autónomo se prevén grandes avances y que la forma en que la sociedad actual se desplaza va a cambiar. Esta nueva tecnología no solo va a producir cambios directos, mejorando los aspectos expuestos en los principios de los MaaS, sino que, además, indirectamente, su introducción va a modificar la situación de muchos otros sectores, como el de los seguros, ¿quién tiene la responsabilidad ante un accidente?, o el de las autoescuelas ¿Se van a necesitar la misma cantidad de permisos de conducción? ¿Y qué pasará con los conductores de transportes de pasajeros como taxis, VTC y autobuses? Es por ello, que la introducción del coche autónomo tiene muchas implicaciones y va a introducir muchos cambios en un futuro próximo.

Históricamente, aunque parezca que los coches autónomos es algo muy novedoso, en realidad esta tecnología lleva mucho tiempo en desarrollo. Ya en 1939, en la exposición mundial de Nueva York, Norman Bel Geddes introdujo un sistema de conducción autónoma, que permitía a un vehículo eléctrico conducir solo, gracias a un circuito eléctrico integrado en el pavimento de la propia carretera. Implementar esta tecnología hubiera implicado tener que rehacer todas las carreteras. El alto costo de este proyecto hizo que no pasara de prototipo. En 1980 Ernst Dickmanns modificó una furgoneta, que, guiada por visión, fue capaz de conducir por calles con el tráfico cortado. Ese mismo año, la *Advanced Research Projects Agency* (DARPA) en busca de nuevos métodos para desarrollar vehículos autónomos, crean un vehículo que utiliza radar láser y visión computarizada para conducirse solo. Posteriormente, en los años 90, Ernst Dickmanns

sigue con distintos proyectos que consiguen recorrer grandes distancias por autopista a gran velocidad. Como ha quedado expuesto, el desarrollo del coche autónomo no es algo novedoso, lo que sí es novedoso es su introducción al público general, despertando el interés de los ciudadanos y por ello el interés de las empresas que ven en los coches autónomos un nicho de mercado, compitiendo entre ellas por liderar el sector. Ahora más que nunca, la investigación y el desarrollo de los coches autónomos está en auge [19].

En la actualidad, grandes marcas como Tesla, Volkswagen o Ford están desarrollando sus propios sistemas autónomos de Nivel 4, que no requieran un conductor, pero ¿qué es esto del nivel?

Para clasificar los coches autónomos según su nivel de autonomía, la Sociedad de Ingenieros Automotrices de los Estados Unidos crea una clasificación basada en 6 niveles. En esta clasificación, que se especifica a continuación, cuanto mayor sea el nivel, mayor es la autonomía del vehículo y, por lo tanto, la independencia del automóvil respecto al conductor [20].

- **Nivel 0 Ninguna autonomía:** En este nivel el conductor no cuenta con ayudas de ningún tipo y por tanto, es el encargado de realizar todas las tareas de conducción.
- **Nivel 1 Asistencia al conductor:** En este nivel el conductor ya si cuenta con ayudas a la conducción, como sistemas que controlan la velocidad del automóvil o la dirección de este. Pero con la peculiaridad de que no se pueden activar al mismo tiempo ambos sistemas, es, por tanto, un nivel de asistencia que facilitan la conducción a las personas, pero que no sustituye al conductor, que sigue teniendo que encargarse del resto de funciones, cayendo por tanto toda la responsabilidad sobre él.
- **Nivel 2 Autonomía parcial:** En este nivel el vehículo utiliza información del entorno en tiempo real y esto le sirve para acelerar, frenar y controlar la dirección por sí mismo, gracias a sistemas como el control de crucero o el cambio de carril automático. Aun así, el conductor sigue teniendo que estar atento en todo momento, supervisando la conducción. Este nivel es el que está presente en la mayoría de los coches autónomos en el mercado actual, como el *Autopilot* de Tesla.
- **Nivel 3 Autonomía condicionada:** En este nivel el vehículo acelera, frena y controla la dirección con los datos que recibe el sistema del entorno. El conductor únicamente tiene que hacerse cargo de la conducción cuando el sistema se lo requiera. Actualmente, se están comercializando los primeros coches en el mercado equipados con sistemas de nivel 3, como el Honda SENSING Elite [21]. También, cabe destacar que la legislación de muchos países no permite la circulación de coches de Nivel 3, dado que el conductor solo debe estar alerta por si el sistema lo requiere para alguna gestión y no tiene que estar atento a la conducción. Aun así, algunos países como Japón [22] o Inglaterra [23] acaban de modificar su legislación, aprobando nuevas regulaciones específicas para este tipo de vehículos.
- **Nivel 4 Alta Autonomía:** En este nivel, los vehículos que incluya este tipo de sistemas serán capaces de conducir por sí mismos, aunque esto queda acotado a zonas en las que el sistema tenga conocimiento. Este nivel se encuentra actualmente en desarrollo y aunque todavía no se comercializan coches con este nivel, se están introduciendo ya los primeros taxis de nivel 4 [24]. En este nivel, el conductor ya no es necesario para manejar el vehículo, dado que no se le va a pedir que tome los mandos del automóvil. Esto es debido a que el coche incluye sistemas secundarios para la conducción para que

en caso de que fallen los primarios, entren en juego, tomando las decisiones necesarias para salvaguardar la seguridad de los ocupantes. Aun así, en ciertas condiciones el automóvil no será capaz de continuar con la conducción autónoma, y, por lo tanto, se mantendrán los equipos necesarios para la conducción manual como el volante o los pedales.

- **Nivel 5 Autonomía total:** Este es el último nivel, donde el vehículo no tiene conductor, solo pasajeros. El vehículo se encargará de todo, las personas únicamente tendrán que introducir el destino al que deseen ir y el coche los llevará hasta él. Los pasajeros podrán dedicarse a cualquier otra tarea, incluso a dormir si así lo desean. Por lo tanto, los equipos necesarios para la conducción manual pueden ser opcionales en este tipo de vehículos. Este nivel todavía se encuentra en un nivel de desarrollo teórico.

Como se ha visto, la conducción autónoma sigue en pleno desarrollo, con cada vez más y más empresas interesadas generando nuevas tecnologías. Pero ¿cómo funcionará esta nueva tecnología, cuando una gran mayoría del público tenga acceso a ella? ¿cómo afectará a la movilidad de las ciudades, cuando un elevado número de coches autónomos este circulando? Para responder a este tipo de cuestiones, será necesario simular todas las posibles soluciones. Entendiendo la simulación como el uso de una representación matemática o informática de un sistema físico con el fin de estudiar los efectos de las restricciones⁴.

Uno de los principales motivos por los cuales se utiliza la simulación basada en agentes (MBA), es el de facilitar el estudio del tráfico de las ciudades. Gracias a estos softwares se pueden identificar los problemas y las soluciones que permitan mejorar el tráfico y descongestionar los núcleos urbanos. La simulación permite probar las distintas ideas y comprobar si estas funcionan o no. Con la introducción del coche autónomo, están surgiendo muchos estudios alrededor de éste, centrados en ver cómo influirá el coche autónomo en el tráfico. Es por ello, por lo que este proyecto se va a centrar en sentar las bases para la simulación, facilitando el trabajo a otros investigadores que quieran estudiar los coches autónomos.

2.4. Agentes

El concepto de agente surge en los años 90 al fusionar diversas disciplinas científicas como la sociología, la psicología, la ingeniería del software, los sistemas distribuidos y la inteligencia artificial. Esta nueva rama comienza con la resolución de problemas de forma distribuida y va evolucionando, entendiendo al agente como una entidad computacional aislada, es el mismo programa el que con las sucesivas ejecuciones va aprendiendo y optimizando sus tareas. La necesidad de tener programas en los que el usuario pueda delegar el trabajo y la toma de decisiones hace que muchos investigadores se interesen por esta nueva disciplina.

Actualmente, la definición del término agente no está claramente fijada, esto es debido a que no hay un consenso en la comunidad científica, el campo de los agentes ha suscitado el interés de un gran número de disciplinas, disciplinas muy dispares que centran sus investigaciones en puntos diferentes, haciendo un uso distinto de esta tecnología. Esto ha provocado que cada campo haya definido el concepto desde una perspectiva distinta y que sea difícil definir este término abarcando a todas las disciplinas a las que implica.

No obstante, una posible definición más general que podría abarcar la definición de muchos campos podría ser la siguiente. Un agente es un sistema informático que es capaz de ejecutarse

⁴ Definición obtenida de Gartner Glossary.

de forma autónoma y flexible ante distintas situaciones, estando ubicado dentro de un entorno en el que debe satisfacer unos objetivos individuales [25].

2.4.1. Arquitecturas de agente

Las arquitecturas especifican el conjunto de módulos de los que se componen los agentes, un conjunto de módulos que interactúan entre si con el fin de cumplir con los objetivos, a su vez esta arquitecturas determinan el conjunto de mecanismos de los que el agente dispone para reaccionar a estímulos, para actuar, para comunicarse con otros agentes, etc.

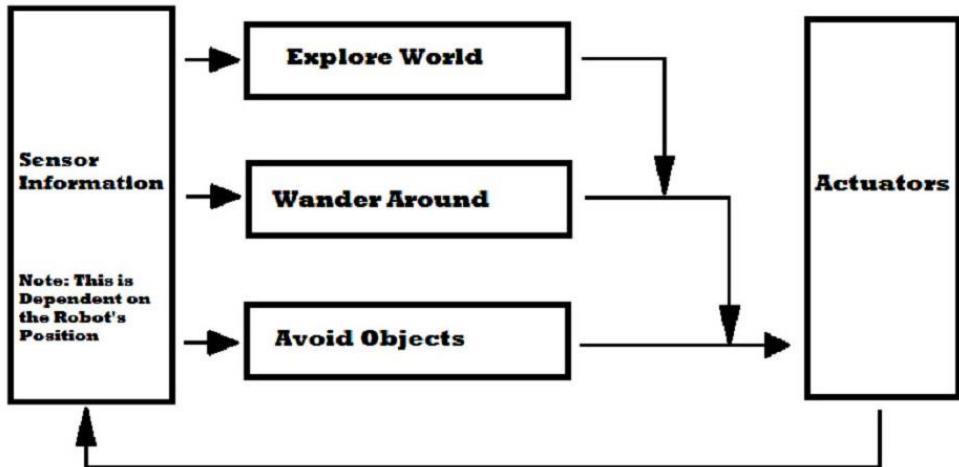


Figura 5. Ejemplo arquitectura subsunción [26]

Existen distintas arquitecturas para los agentes, es decir, distintas formas por las que se puede subdividir el trabajo, dividiendo las tareas en acciones. Las arquitecturas también difieren en la forma en que se planifica la ejecución de estas acciones con el fin de satisfacer el propósito del sistema. Las principales arquitecturas de agentes son las siguientes:

- **Arquitecturas Deliberativas:** Tiene como elemento central de razonamiento un modelo simbólico. El sistema parte de un estado inicial que contiene un conjunto de planes y un estado objetivo a cumplir. Para satisfacer con su propósito, el sistema planifica los pasos que se han de dar.
- **Arquitecturas Reactivas:** No tiene como elemento central de razonamiento un modelo simbólico y no hace uso del razonamiento simbólico complejo, con esto se evitan los problemas asociados a la representación simbólica del conocimiento. Las arquitecturas de subsunción son un ejemplo de estas arquitecturas y se caracterizan por trabajar con jerarquías de tareas, que definen como actuara el agente frente a los estímulos.
- **Arquitecturas Híbridas:** Combina las arquitecturas deliberativa y reactiva. El agente se compone como mínimo de dos subsistemas, uno por cada arquitectura. Por ello, esta arquitectura se divide en capas o niveles, donde, o todos tienen acceso a los sensores y actuadores (horizontal) o donde solo uno tiene el acceso a estos (vertical). Usualmente estas arquitecturas se estructuran en tres capas:
 - o Nivel bajo: El primer nivel es el reactivo, dependiendo de los estímulos que se van recibiendo se toman las decisiones
 - o Nivel intermedio: El segundo nivel es el de conocimiento, enfocado en la información que se tiene del medio.

- o Nivel alto: El tercer nivel es el social, este se ocupa de aspectos sociales del entorno, como la información de otros agentes, las intenciones, los deseos, etc.

Aunque los agentes se pueden programar con lenguajes generales orientados a objetos como java o C++, también existen lenguajes específicos para programarlos como Agent0 o PLACA. Por otro lado, están las plataformas de agentes, *middlewares* que actúan como entorno de ejecución para los agentes, apoyando el despliegue y la gestión de estos, proporcionando a los agentes una serie de servicios.

2.4.2. Propiedades de los agentes

Un agente percibe su entorno mediante sensores, los datos obtenidos son procesados por el agente para conocer la situación actual del entorno. Con la información de que disponible el agente, este debe decidir qué acciones realizar con el fin de llegar a cumplir con sus objetivos individuales.

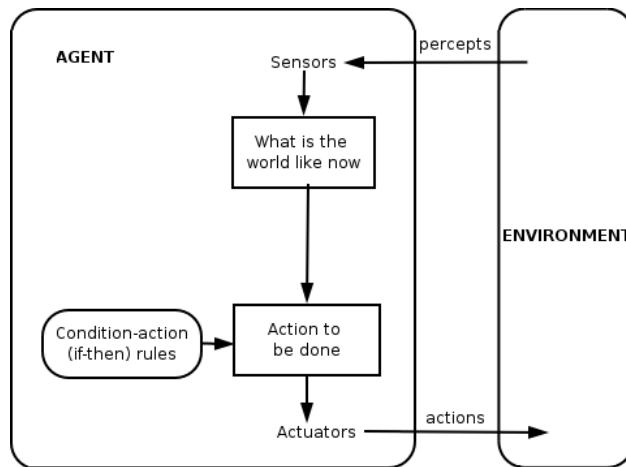


Figura 6. Interacción de los agentes [27]

Las principales propiedades que caracterizan a los agentes son [25]:

- **Autonomía:** Los agentes deben ser capaces de actuar por sí mismos sin que un usuario u otro agente tenga que intervenir.
- **Sociabilidad:** Los agentes deben ser capaces de interactuar entre ellos con una cooperación explícita que se da mediante la comunicación, gracias a la cual se intercambian información.
- **Reactividad:** Los agentes deben ser capaces de reaccionar ante cambios imprevistos que se producen en el entorno, estímulos a los que tendrán que responder con celeridad para no quedarse bloqueados.
- **Iniciativa:** Los agentes deben ser capaces de actuar por si solos para cumplir con sus objetivos, por ello, además de reaccionar a los estímulos, los agentes deben tomar la iniciativa y tomar sus propias decisiones.

2.4.3. Nuevas tecnologías a partir de los agentes

La inteligencia artificial, se ocupa de la creación de sistemas computacionales aislados que resuelven problemas utilizando distintos métodos, estos métodos se basan en el conocimiento, en el aprendizaje o en algoritmos de búsqueda. Pese a las innovaciones que han aportado los agentes a la inteligencia artificial, como la autonomía y el aprendizaje, existe una limitación para estos sistemas, asociada al hecho de que éstos trabajan solos, sin interacción alguna con otros sistemas.

El avance de la informática introdujo la concurrencia y la distribución de recursos entre sistemas, surgiendo la inteligencia artificial distribuida que trabaja con más de un sistema a la vez, coordinándolos y trabajando de forma semiautónoma, superando gracias a esta coordinación, las limitaciones de trabajar individualmente.

Con la aplicación de los avances de la inteligencia artificial distribuida al campo de los agentes, surgen nuevos sistemas que mantienen la idea de la autonomía y el aprendizaje, pero que pasan de utilizar un solo agente, a un conjunto de agentes que trabajan juntos para conseguir una meta. Esta puerta abre el desarrollo a nuevas ramas como los sistemas multiagente, sistemas basados en agentes, o el modelado basado en agentes, entre muchos otros. Al igual que con la definición del término de agente, el interés por estas nuevas ramas hace que se expandan rápidamente y que sea difícil esclarecer a priori las diferencias entre estos nuevos sistemas.

2.5. Sistemas multiagente (MAS)

Los sistemas multiagente (MAS) son sistemas computarizados compuestos por subsistemas menores, es decir, un conjunto de agentes inteligentes que interactúan entre sí. Los MAS se basan en la coordinación de los subsistemas de los que se componen, haciendo que los objetivos individuales de cada subsistema se integren en un objetivo común.

2.5.1. Características

Los sistemas multiagente están compuestos por una colección de agentes, cada uno de ellos definido por sus propias habilidades: adquisición de datos, comunicación, planificación y actuación.

El conjunto de los agentes tiene asignado una misión, para cumplirla, estos tendrán que trabajar juntos coordinándose, cooperando, o negociando entre ellos. Esta misión puede dividirse en diferentes tareas independientes de modo que puedan ejecutarse en paralelo. No existe un sistema global de control, los datos están descentralizados y la computación es asíncrona. Por ello, es el propio conjunto de agentes el que se asignará las tareas más apropiadas a cada uno de ellos, teniendo en cuenta la adecuación de un agente a una tarea y a la consecución de la misión común. De modo que los agentes deciden dinámicamente qué tareas realizan.

Un agente tiene un conocimiento incompleto, esta limitación puede ser tanto del conocimiento del entorno, como de la misión del grupo, o como de las intenciones de los demás agentes. Por lo tanto, un agente será más adecuado para realizar unas tareas u otras en función de los conocimientos que tiene, en función de si posee la habilidad requerida y en función de sus capacidades de procesamiento.

2.5.2. Ventajas de los MAS

Algunas de las ventajas que aportan los sistemas multiagente al distribuir las decisiones entre los agentes son las siguientes:

- La autonomía de cada agente del sistema, dado que los agentes no dependen de la condición de otros agentes para escoger las tareas. Cada uno decide las tareas más relevantes en función de su entorno.
- No es necesario que toda la información del MAS esté en un único agente, dado que las decisiones internas de cada agente se toman individualmente a partir de los datos internos de los que dispone.
- La información del sistema se transfiere a un hipotético sistema central ya que algunos agentes pueden tomar decisiones que involucran a todos los demás. Por ello, el proceso de comunicación es de vital importancia, pues permite a los agentes cooperar entre sí.
- El sistema es muy robusto, puesto que, al tener la información en distintos puntos, en el caso de que un agente falle, el sistema puede seguir funcionando.

2.5.3. Propiedades

Como se ha comentado, los sistemas multiagente son sistemas complejos compuestos por muchas variables. Para poder simular y conseguir buenos resultados estos sistemas exigen al usuario un trabajo previo de preparación y configuración.

- El modelo computacional: Nivel de distribución que se alcanza a la hora de llevar a cabo los cómputos, este define hasta qué punto se puede construir un modelo que permite distribuir el conocimiento.
- El nivel de autonomía de los agentes: nivel de autonomía que pueden tener cada uno de los agentes en el sistema, este establece hasta qué punto pueden decidir por sí mismos si ejecutan una u otra acción.
- Los recursos de los agentes disponibles en el sistema: referido a si los recursos de los que disponen los agentes son limitados y compartidos o si, por el contrario, cada agente tiene sus propios recursos y estos son ilimitados.
- El nivel de interacción entre los agentes: la complejidad en las interacciones variará, permitiendo o no el intercambio de información a alto nivel.

2.5.3.1. Competencias

La descripción de competencias establece la subdivisión del problema en tareas, subtareas y sus relaciones. Determina, por tanto, como se distribuye el problema entre los diferentes agentes y como son las interacciones entre ellos. Esta asignación de tareas se puede establecer durante la definición del problema de forma determinista, o de forma dinámica durante la ejecución, teniendo en cuenta cómo va evolucionando el sistema.

2.5.3.2. Modelos de agentes conocidos

Los modelos de agentes conocidos sirven para mostrar a los agentes de forma local la información que tiene ese agente de otros. Modelos que contienen las características, capacidades, conocimientos y creencias de los agentes externos, que le sirven para saber comunicarse, para intentar predecir el comportamiento de otros agentes, para planear y coordinarse. Utilizando los modelos de agentes se consigue que el sistema sea dinámicamente autoconfigurable, es decir, que el sistema se pueda adaptar a cambios en el entorno o a nuevas configuraciones del grupo, permitiendo la inclusión o exclusión de un agente en un grupo.

2.5.3.3. Comunicación entre agentes

La comunicación entre los agentes permite la coordinación y el intercambio de información. Esta comunicación se produce a través de los canales de comunicación, mediante un protocolo que puede tener diversos niveles de complejidad, pudiendo darse mediante una sintaxis estructurada o mediante conversaciones no estructuradas.

2.5.3.4. Comportamiento

El comportamiento coherente de los agentes es de vital importancia para el sistema. La coherencia se califica mediante la calidad y claridad de la solución final, más los fallos no críticos del sistema, es decir, los fallos producidos por agentes, que no son determinantes en la solución final. Una de las variables que influyen en el comportamiento de un sistema es el tipo de organización de los agentes, los diferentes tipos de organización afectan de distinta forma a la coherencia del sistema:

- Organización centralizada: Solo un agente toma las decisiones.
- Organización tipo mercado: Mediante ofertas y contratos se resuelven las interacciones entre agentes.
- Organización de comunidad plural: Las decisiones de cada agente son optimizadas por el resto.
- Organización con reglas de comportamiento: Un grupo de agentes expertos controlan las interacciones mediante preguntas y respuestas.

Otros factores que repercuten en la coherencia son: la sincronización de acciones por medio de una planificación global, la reducción de los puntos de interacción, la propia comunicación, la mejora de los modelos de agentes conocidos.

2.5.3.5. Puntos de interacción

La necesidad de comunicación de un agente con otros agentes de un grupo se cubre mediante llamadas a funciones de petición. Los puntos de la aplicación donde se da esta comunicación se llaman puntos de interacción.

2.5.4. Aplicaciones

Los sistemas multiagente están siendo utilizados en campos muy diversos como el transporte y la logística [28], la robótica [29], la geografía [30], sistemas de defensa coordinados [31], sistemas de fabricación [32], entre otros. Esto es debido a que estos sistemas son de gran utilidad cuando:

- Los problemas no pueden ser resueltos por un solo agente individual
- Existen varios métodos para resolver un problema
- Los problemas son físicamente distribuidos
- Se posee un sistema distribuido donde los componentes hardware están unidos mediante una red.
- La complejidad de un estudio es muy alta y se necesitan varios puntos de vista distintos para plantear la solución.
- El sistema tiene que adaptarse a cambios en el entorno

2.6. Simulación basada en agentes

La simulación basada en agentes también conocida como Modelado Basado en Agente (MBA), es un modelo computacional dirigido a la simulación con agentes, los MBA están muy relacionados con los sistemas multiagente y surgen a partir de ellos. La diferencia entre los sistemas multiagente y la simulación basada en agentes radica en que estos últimos, tienen por objetivo comprender y explicar el porqué del comportamiento colectivo de los agentes, utilizando para ello agentes que no necesariamente deben ser inteligentes, pues obedecen a reglas simples. Estos sistemas se suelen utilizar para estudiar sistemas naturales que contienen irracionalidad, instintos, y otros comportamientos. Mientras que lo MAS se centra en la resolución de problemas prácticos, en el diseño de los agentes o en dar solución a problemas específicos de la ingeniería. La simulación basada en agentes ha conseguido que los avances en ingeniería que se han conseguido mediante los MAS, puedan trasladarse con gran éxito a otros campos muy diversos como pueden ser las ciencias sociales o la biología [33].

Este modelo computacional permite estudiar la complejidad y la no-linealidad típica de muchos fenómenos sociales mediante la simulación con agentes, pudiendo recrear y comprender como funciona un sistema y entender las razones por las que se dan distintos fenómenos complejos, además de poder llegar a predecirlos. Es por ello, por lo que los MBA son considerados como un tipo de modelo a microescala. Donde la simulación se compone de un conjunto de reglas simples y de un conjunto de agentes, que con distintas conductas individuales (microescala) interactúan entre ellos dentro de una red o entorno, generando cambios a un nivel mayor, el de todo el sistema (macroescala) [34].

Los agentes son las unidades elementales que forman la población de un MBA, estos agentes pueden ser tanto entidades individuales, como puede ser una persona, o como entidades colectivas, como puede ser un hogar. Estos agentes tienen que cumplir con una serie de objetivos, respetando las reglas impuestas. En el caso de este trabajo se van a utilizar los MBA para la simulación de la circulación de personas en una ciudad. Ya que estas simulaciones son

de gran utilidad para el estudio de nuevos planes urbanísticos, para la mejora de las vías existentes en una ciudad, o para la introducción de nuevos medios de transporte entre otros.

Vistas las diferencias entre MAS y MBA, cabe comentar que las arquitecturas, las propiedades, la forma en que se comunican los agentes y demás apartados vistos dentro del apartado 2.5 coinciden en ambos tipos de sistemas.

2.7. MATSIM

Matsim como su nombre indica, Multi-Agent Transport Simulation, es un entorno de trabajo (*framework*) de código abierto, para implementar simulaciones de transportes a gran escala basadas en agentes.

2.7.1. Historia y Bases

Matsim es creado inicialmente por Kai Nagel, que después de trabajar en varios proyectos de análisis de los sistemas de transporte como *TRansportation ANalysis SIMulation System* (TRANSIMS) y otros estudios en torno al modelado de tráfico, decide crear un programa de código abierto a partir de sus conocimientos utilizando C++ como lenguaje de programación. Tras la revisión del proyecto que hace Michael Balmer, ambos investigadores se unen decidiendo continuar el desarrollo del programa en Java, debido a que este lenguaje aportaba grandes ventajas que no disponía C++, utilizando como base el código escrito por Balmer en Java. El interés que suscitó este proyecto hizo que nuevos investigadores se unieran a él, implementándose sucesivas mejoras hasta llegar al simulador actual. Este programa surge de la fusión de cuatro corrientes de investigación:

- **Modelado microscópico del flujo de tráfico:** fue la base para la teoría del flujo del tráfico, pero tenía la gran limitación de trabajar con links (calles) individuales o con pequeñas secuencias de links. El aumento de la capacidad computacional de los ordenadores junto al avance en los modelos de flujo ha permitido que esta rama se desarrolle. Creándose nuevos modelos de flujo de tráfico, necesarios para poder representar las interacciones de los agentes con la red viaria y de los agentes entre sí.
- **Física computacional:** Dado que Matsim está ideado para ejecutar grandes simulaciones, la optimización y paralelización del trabajo resulta de vital importancia. Los aportes de esta rama de la física han hecho posible que Matsim pudiera adoptar modelos simples y rápidos que han conseguido una capacidad de procesamiento mucho mayor con simulaciones de más de 10^8 agentes.
- **Modelado basado en agentes:** Base principal de Matsim, que utiliza el concepto de agentes para representar a las personas que se desplazan por la red, percibiendo su entorno y respondiendo a eventos que les surjan. Se genera así un escenario gracias al modelado basado en agentes, que luego Matsim va a utilizar en clave global para buscar soluciones a problemas, entender el funcionamiento de la red preexistente permite su mejora.
- **Algoritmos coevolutivos:** Desde las primeras aproximaciones para la asignación de transportes basadas en el equilibrio de Nash, se han ido introduciendo sucesivas mejoras como el aprendizaje a partir del contexto o técnicas para integrar datos de iteración en iteración. Estas mejoras han supuesto un salto cualitativo de esta aproximación inicial hacia los esquemas de búsqueda coevolutiva, que consiguen

obtener la planificación óptima para cada agente individualmente. Estos métodos heurísticos parten de los algoritmos evolutivos, pero no se centran en el *fitness* global, si no que la evaluación se basa en las interacciones entre los individuos, ideas que se han aportado a los agentes para desarrollar Matsim.

Matsim simula escenarios. Un escenario está compuesto por una población de agentes y unos planes iniciales, ambos basados en personas y desplazamientos reales que se producen en el área estudiada. Estos escenarios también incluyen las localizaciones reales de edificios, como puede ser los hogares o los lugares de trabajo y la red viaria, en donde los agentes competirán entre sí para cumplir satisfactoriamente sus planes, modificando para ello sus comportamientos hasta llegar a un equilibrio.

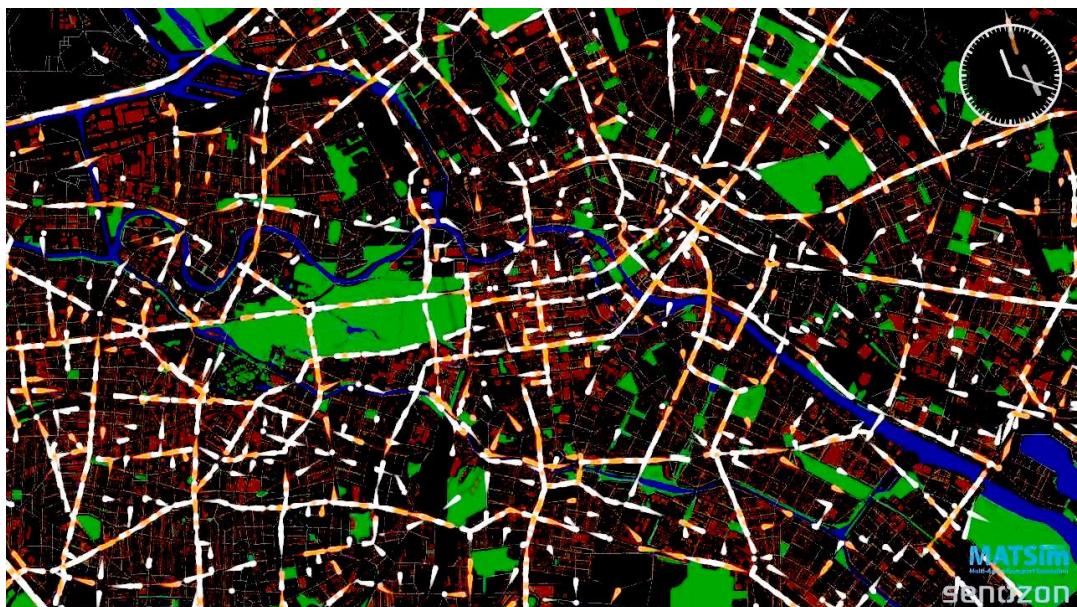


Figura 7. Visualización de la simulación en Matsim [35]

Este software está diseñado para ejecutar simulaciones de 24 horas, aunque se podría implementar escenarios de mayor duración. Matsim ofrece al usuario la posibilidad de jugar con estos escenarios y experimentar con los parámetros de la función de utilidad, con el porcentaje de población usado en la simulación, experimentar también con los parámetros del algoritmo, con los que se pueden modificar variables, como el porcentaje de agentes que puedan replanificar sus planes en las distintas iteraciones, o el comportamiento de los agentes, etc.

2.7.2. Algoritmo Coevolutivo

Un algoritmo evolutivo conseguiría una optimización global, lo que equivaldría a que cada persona reflexionara globalmente su movilidad, pensando que desplazamientos suyos podrían mejorar la circulación de la ciudad. Esta solución no es realista, ya que las personas buscan mejorar únicamente sus desplazamientos, intentando reducir el tiempo de viaje al mínimo para llegar antes a sus destinos. Por ello, Matsim no utiliza los algoritmos evolutivos, para busca el equilibrio Matsim emplea algoritmos coevolutivos, donde los agentes compite entre sí por el tiempo y espacio, con el fin de optimizar sus planes diarios iteración a iteración.

Para optimizar la cadena de actividades diarias, los agentes pueden modificar la ruta seguida al destino, la hora de salida, el tipo de transporte utilizado e incluso el destino al que se dirigen para realizar la actividad. Cada agente tiene un espacio de memoria donde guarda los planes que posee, tanto los iniciales, como los que se van generando con las iteraciones, cada uno de

estos planes contiene una serie de actividades diarias y una puntuación asociada. En caso de que un agente tenga demasiados planes, aquellos que tengan una puntuación menor se borran.

2.7.3. Esquema de funcionamiento

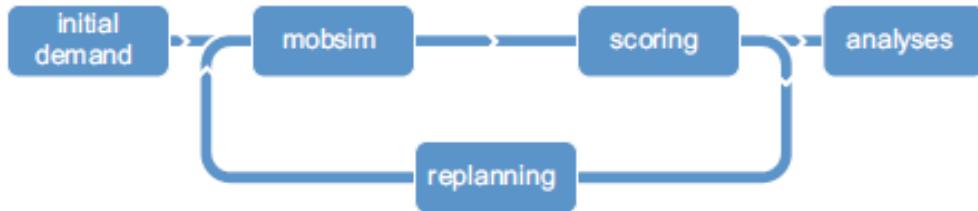


Figura 8. Bucle funcionamiento de Matsim [36]

Una vez que el usuario haya seleccionado las múltiples opciones configurables de la simulación, Matsim funcionará siguiendo el esquema mostrado en la Figura 8.

En el primer paso se introducen los planes iniciales para que a continuación cada agente seleccione un plan de su memoria, sin tener en cuenta las puntuaciones de estos. A continuación, se comienza con la primera iteración del programa simulando la movilidad de los agentes por el mapa durante 24 horas (mobsim), terminada la ejecución, el sistema calcula las puntuaciones de los planes seleccionados por los agentes en base a sus actuaciones. Posteriormente, se pasa al módulo de replanificación, en el cual, un porcentaje predefinido de los agentes clona el plan que tenían seleccionado y lo modifica, cambiando las variables anteriormente mencionadas: los horarios de partida, las rutas, el modo o el destino, para luego añadir este antiguo plan modificado a su espacio de memoria como un nuevo plan. Estas modificaciones varían en función del módulo de replanificación, el cual puede estar configurado para realizar mutaciones aleatorias o para realizar modificaciones informadas que optimicen el plan, finalizando así con la primera iteración. Este proceso relatado compuesto de simulación, puntuación y replanificación se va a seguir realizando hasta que se hayan realizado todas las iteraciones definidas o hasta que se estabilice la puntuación media de la población, es decir, cuando continuar con más iteraciones no vaya a mejorar los resultados cualitativamente. Terminadas las iteraciones se pasa al módulo de análisis, donde se pueden ver los resultados obtenidos.

Matsim tiene un diseño modular, por ello, el funcionamiento básico descrito en este apartado se puede modificar agregando distintos módulos, esto permite añadir nuevas funcionalidades que no posee el esquema básico inicial, que le añaden una mayor complejidad

2.7.4. Modelo de Flujo del Trafico

El modelo de flujo de tráfico predeterminado es QSim aunque existen muchos más. Este modelo discretiza las calles en celdas, modela el tráfico a través de colas, obtiene las velocidades de los automóviles a partir de agregados y estudia el tráfico como un flujo de coches en vez de observar cada vehículo individualmente.

En QSim el hecho de que un vehículo entre en un *link*, es decir, en una calle, se representa mediante una cola a la que este nuevo vehículo se agrega en última posición. El vehículo va a estar dentro de esta cola un periodo correspondiente a la suma del tiempo estimado que tarda el vehículo en recorrer la calle, más el tiempo de espera hasta convertirse en el primer elemento de la cola, más el tiempo en el que el siguiente *link* al que se desea ir este libre y deje entrar al vehículo.

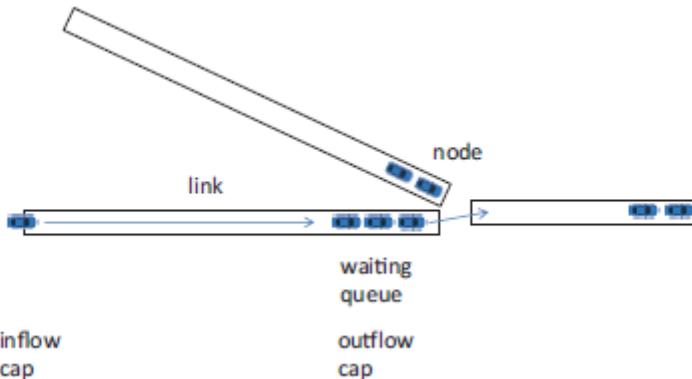


Figura 9. Modelo de flujo del tráfico [36]

Este modelo de flujo basado en espacios de tiempo para los agentes es muy eficiente, pero simplifica variables que afectan a la circulación. Por esta razón, se han llegado a crear otros modelos como JDEQSim que combina el sistema de colas con un sistema de actualización basado en eventos, donde no existe un proceso de actualización basado en el tiempo, el sistema solo interactúa con los agentes cuando se requiere una acción. Un modelo de flujo de tráfico se basa en dos atributos de los enlaces:

- Capacidad de almacenamiento: Fija la cantidad de vehículos que puede haber en un *link*.
- Capacidad de flujo: Fija la cantidad de vehículos que pueden salir del *link* por unidad de tiempo.

QSim no tiene fijada una cantidad máxima de vehículos que pueden entrar en un *link*, lo que puede generar atascos. También se pueden agregar módulos que agreguen señales o módulos con los que se puedan modelizar múltiples carriles. Cada modelo de flujo tiene sus beneficios y sus desventajas, siendo tarea del usuario estudiar detalladamente que modelo le interesa más.

Para entender Matsim en profundidad los desarrolladores del simulador han escrito un libro que contiene toda la información respecto al programa. Explicando su funcionamiento, su configuración para poder empezar a simular, ejemplos de proyectos basados en Matsim, etc. [36].

3. Diseño de la solución

Los sistemas basados en agentes que simulan el tráfico necesitan una población y una red viaria para poder simular, el problema de los simuladores existentes reside en que estos no aportan ni la población, ni la red vial por la cual se mueven los agentes. Si además se quiere realizar un estudio preciso, no valdrán poblaciones de muestra que contengan pocos individuos, o que los individuos introducidos no sean representativos de la población real a la que se pretende simular. Por lo tanto, la población y la red deben ser específicos para cada estudio, cuanto más fiel a la realidad sean los datos introducidos como entrada, mejores resultados se podrán obtener con la simulación. Esto hace que el trabajo previo a la simulación pueda requerir mucha más carga de la esperada.

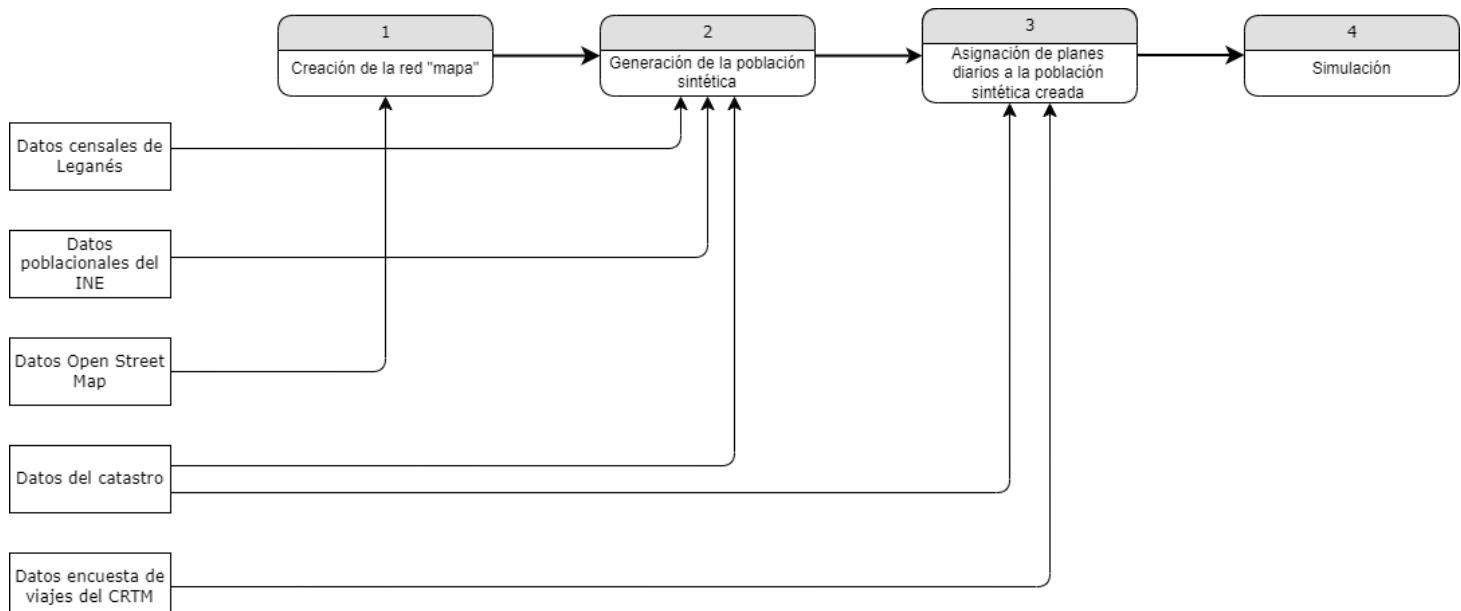


Figura 10. Mapa de los procesos a implementar para poder simular

En la Figura 10 se muestra un esquema simplificado del proceso de desarrollo que se va a seguir para poder llegar a simular. El esquema consta de tres grandes apartados a realizar antes de la simulación y las fuentes de datos utilizadas para cada uno de ellos.

La metodología de desarrollo ágil utilizada para el desarrollo de este trabajo es Scrum, ya que se deseaba seguir una estrategia de desarrollo incremental, dividiendo el desarrollo en pequeños proyectos que tras su consecución den acceso a la simulación de transportes. Durante el tiempo de desarrollo se han ido solapando las distintas fases, introduciendo distintas mejoras y revisando los apartados.

Scrums hace una diferenciación de roles entre los miembros del equipo, en el caso de este trabajo, el tutor del TFG asumirá los roles de *product owner* y de *scrum manager*, dado que es a la vez el cliente que demanda el proyecto y el director que supervisa el trabajo. Mientras que el estudiante asumirá el rol de *development team* ya que se encargará de diseñar, implementar y testar el proyecto.

3.1. Generación de la red

En el primer paso se va a generar la red viaria de Leganés. Para ello, se utilizarán los datos obtenidos de OpenStreetMap mediante el programa JOSM. Este software fue elegido porque

contiene una extensión específica para Matsim creada Andreas Neumann y Michael Zilske [37], la cual facilita la obtención de los datos necesarios para simular, al dar como output un fichero con la estructura que requiere Matsim.

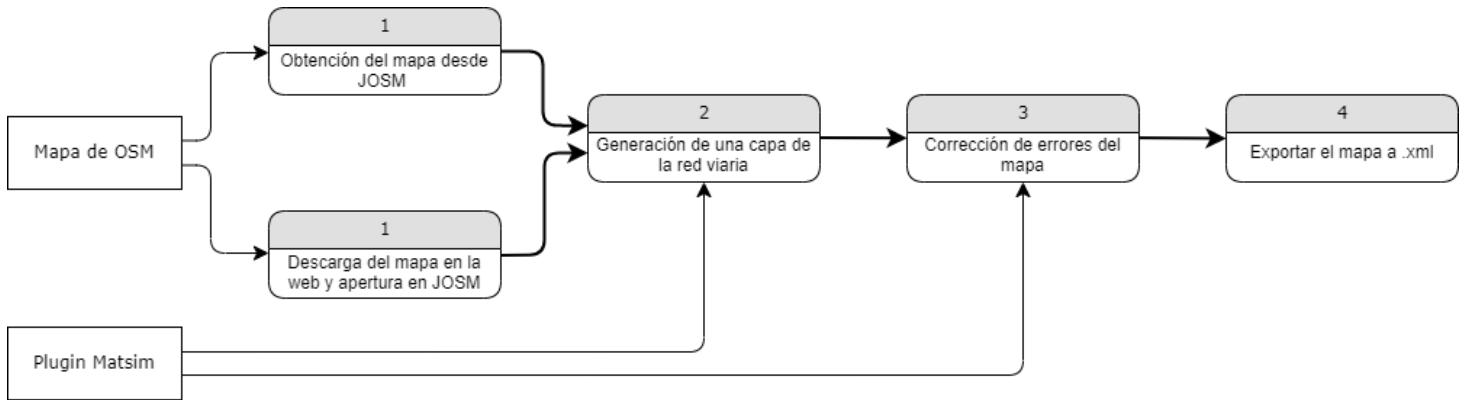


Figura 11. Mapa de procesos a implementar para obtener la red

En el esquema mostrado en la Figura 11 se muestran los procesos a realizar para poder obtener la red en la cual se va a simular. En el primer paso se obtienen el mapa de OpenStreetMap, para luego generar una capa que únicamente contiene la red viaria, mediante la extensión de Matsim, para luego en el tercer paso corregir los posibles errores y mejorar la red, para en el último paso obtener el archivo XML resultado al exportar la capa de la red.

3.1.1. OpenStreetMap

OpenStreetMap (OSM) es un proyecto colaborativo, donde gracias a los aportes de numerosos usuarios, se ha generado un mapa de toda la tierra, aunque dependiendo de la región el nivel de detalle varía. Este mapa, de gran exactitud, contiene una gran cantidad de datos como la delimitación de ciudades o barrios, los edificios clasificados por tipos o las redes viarias, que a su vez contienen datos como el límite de velocidad de la vía o el número de carriles. Esta herramienta es de uso libre bajo una licencia abierta, es fácil de utilizar y permite a los usuarios descargar los mapas de cualquier zona del mundo.

Para obtener datos de OSM existen herramientas con las que se pueden hacer consultas, Overpass turbo [38] es una de estas herramientas y permite realizar consultas a los datos de OSM haciendo uso del lenguaje *Overpass Query Language* (Overpass QL). A parte del código obtenido como resultado de las consultas, se pueden visualizar los resultados en un mapa interactivo, ayudando al usuario a comprobar si las consultas son correctas, lo que simplifica el proceso. El mayor problema de OSM viene de que esta herramienta muchas veces no está actualizada, ya que depende de la colaboración desinteresada.

3.1.2. JOSM

Java OpenStreetMap (JOSM) es una herramienta muy útil para poder trabajar con OpenStreetMap. Este software diseñado para escritorio ha sido creado para editar los datos de los mapas de OpenStreetMap y actualmente posee una gran cantidad de herramientas avanzadas, que facilitan la actualización y ampliación del mapa.

JOSM incluye una funcionalidad que permite al usuario descargar desde el servidor de OSM un área seleccionada directamente desde el mapa, lo que facilita trabajar con esta herramienta. Sin

embargo, JOSM limita el tamaño del mapa a descargar, impidiéndolo cuando se excede un determinado volumen de datos que posee el área seleccionada a descargar. En áreas urbanas donde la cantidad de datos es mucho mayor, el área que este programa permite descargar es menor, teniendo que reducir el área seleccionada si se desea descargar el mapa, no obstante, en áreas rurales donde existen menos datos, se amplían los tamaños del área a descargable.

En el caso de que el mapa requerido tenga unas dimensiones que no admite JOSM, caben dos soluciones. Si el mapa que se necesita no es mucho más grande que el área que JOSM permite descargar, se pueden hacer varias descargas y es el propio programa el que acopla las nuevas zonas descargadas sin que el usuario tenga que hacer nada. En el caso de Leganés que mide 43,25 km² solo fueron necesarias dos descargas para obtener el mapa completo. No obstante, si se desea estudiar una zona mucho más amplia, como grandes ciudades o regiones enteras, esto va a implicar que se tengan que realizar muchas descargas y por ello, este método no sería el más indicado. La segunda opción consiste en utilizar páginas web como bbbike.org [39], que contienen archivos muy grandes para descargar con la extensión “.osm”, para posteriormente abrirlos en JOSM. Aunque una de las páginas más completas para descargar estos mapas de gran tamaño es Planet.osm [40], dado que está vinculada a OSM. Planet.osm posee un archivo con todo el mapa del mundo, además de extractos de este archivo principal con mapas de países y regiones específicas.



Figura 12. Mapa de Leganés descargado de OpenStreetMap [41]

En la Figura 12 se puede observar el mapa de Leganés en bruto tras ser descargado con todos los datos que ofrece OpenStreetMap. Dado que el fichero *network* de Matsim solo contiene información de la red viaria, la mayoría de los datos que ofrece OSM son innecesarios para este apartado.

Utilizando el *plugin* de Matsim para JOSM se puede eliminar con facilidad la información innecesaria, quedando solo representada la red viaria. Sin embargo, antes de utilizar dicho *plugin*, es necesario modificar el formato de las coordenadas, ya que si esto no se hace, las

coordenadas que se introducirían en el fichero de salida estarían en formato WGS84, formato no deseado tal y como se cita textualmente en la guía de Matsim [36] “*In several input files, you need to specify coordinates, e.g., for network nodes. For the time being, we strongly advise not to use WGS84 coordinates, or any other spherical coordinates*”⁵. Esta recomendación de Matsim de no usar coordenadas esféricas, viene dada por el constante cálculo de distancias entre dos puntos durante la simulación. Dado que el cálculo de distancias en coordenadas esféricas es mucho más complejo y costoso computacionalmente, realizar todos estos cálculos con coordenadas esféricas podría ralentizar el programa. Por esta razón, Matsim utiliza coordenadas cartesianas para así poder utilizar el teorema de Pitágoras en el cálculo de la distancia entre puntos, simplificando los cálculos.

JOSM tiene una pestaña denominada proyección del mapa dentro de preferencias en donde se puede modificar el tipo de coordenadas que se quieren utilizar, aquí se puede escoger UTM como método de proyección, teniendo que indicar el hemisferio y la zona UTM.

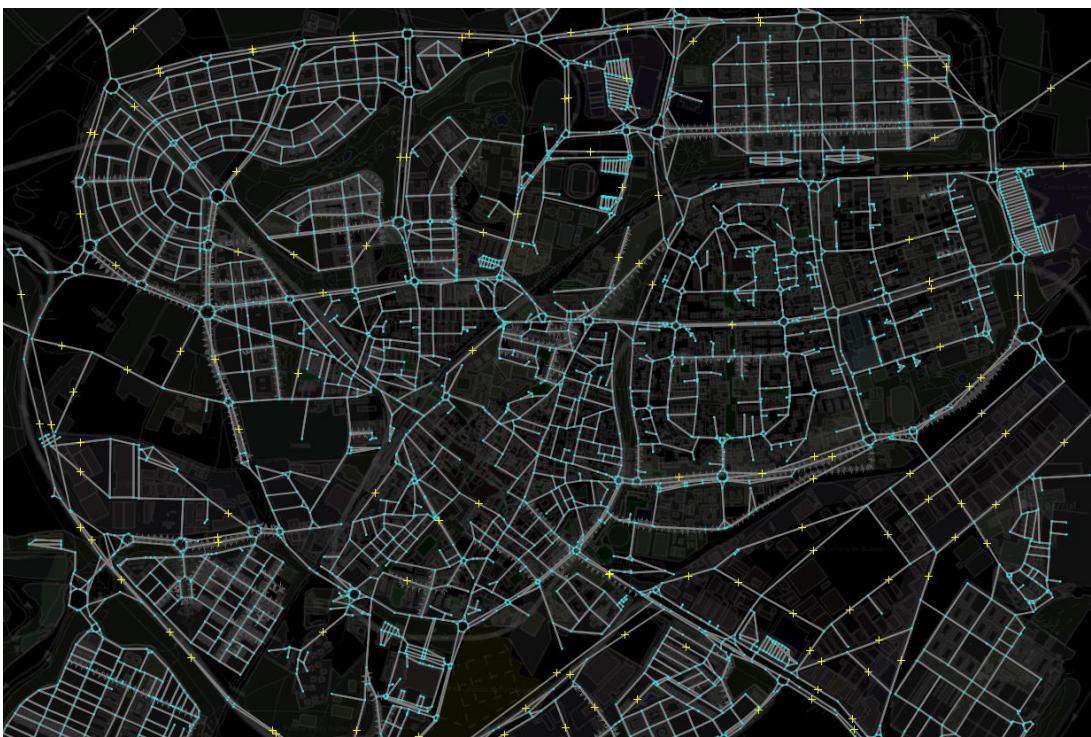


Figura 13. Mapa de la red viaria de Leganés obtenida con el *plugin* [42]

Una vez seleccionas las coordenadas UTM solo queda hacer uso del *plugin* de Matsim, generando una nueva capa que únicamente contendrá la red viaria de la región, tal y como se muestra en la Figura 13 [43]. A su vez, este *plugin* contiene nuevas funcionalidades, como una tabla en donde se pueden consultar las siguientes variables de las vías seleccionadas:

- Id
- Longitud de la calle
- Número de carriles
- Límite de velocidad
- Capacidad

⁵ Fragmento del libro *The Multi-Agent Transport Simulation MATSim* [36]

Esta tabla que se muestra en la **Figura 14**, permite al usuario ordenar los datos por cualquiera de las variables mencionadas, pudiendo identificar datos específicos con facilidad.

Links: 7.109 / Nodes: 3.267					
id	length	freespeed	capacity	permlanes	
32574066_1_r	97,196	22,222	2.000	1	
30978832_0	117,524	4,167	300	1	
121842219_0	184,235	16,667	1.000	1	
28669379_6_r	16,103	8,333	600	1	
421145313_0	22,765	16,667	1.000	1	
889921786_0	17,777	11,111	1.500	1	
28048157_4	65,582	11,111	600	1	

Figura 14. Tabla del *plugin* con datos de las vías [42]

Antes de exportar el mapa a un XML apto para Matsim, se debe revisar que los datos aportados por el mapa coincidan con la realidad, dado que OSM es un mapa colaborativo en el que cualquiera puede trabajar. Se deberían, por tanto, revisar y corregir los datos relativos a los límites de velocidad, el número de carriles, el tipo de vía o que alguna calle no esté representada en el mapa.

Una funcionalidad del *plugin* que puede ayudar al usuario a corregir los errores del mapa es un sistema de validación que funciona al exportar el archivo, esta funcionalidad muestra la existencia de calles duplicadas, *id* duplicados o campos de calles en *null*, permitiendo así corregir con facilidad los posibles errores. En el caso de este trabajo, aparte del uso del sistema de validación, también se han corregido otros aspectos, como obviar las pequeñas calles que aparecen en la Figura 13, que representan entradas a garajes que solo entorpecerían la simulación. También se pueden apreciar los parkings al aire libre de los centros comerciales, de los que se han eliminado las múltiples calles en paralelo dado que estas serían reconocidas como calles normales y podrían influir negativamente en la simulación.

Finalmente, tras las modificaciones pertinentes, para exportar el mapa, simplemente hay que guardar el archivo y elegir la extensión XML que usa Matsim, obteniendo así el fichero que va a utilizar el programa como red en la simulación.

3.2. Generación de la población

Como ya se ha comentado, conseguir una población representativa de lo que se vaya a simular, es de vital importancia. Debido a que los simuladores no incluyen la población con la que simular, es el usuario el que debe obtener los datos de las personas antes de iniciar las simulaciones. Sin embargo, obtener los datos completos de una población no es factible dado que los estados ofrecen datos incompletos por temas de protección de datos, dejando en abierto datos que no contienen todas las variables u ofreciendo una pequeña muestra de la población censada. Otro problema puede ser que los datos no sean actuales, puesto que el censo se suele realizar cada 10 años.

Dado que no es posible conseguir la población completa, se han desarrollado una serie de métodos y algoritmos que parten de una pequeña porción de datos reales y restricciones obtenidas de distintas fuentes y que, mediante la correlación en estos datos, se generan las entradas de los modelos basados en agentes, es decir, la población sintética realista de la región que se quiere simular. Estos programas se denominan **sintetizadores de población** y cada uno utiliza distintos métodos para lograr la población objetivo.

La generación de la población se desarrolla en dos pasos, un primer paso consiste en crear a las personas que serán los futuros agentes de la simulación, y un segundo paso en el que se asignan planes diarios a las personas previamente creadas, sin embargo, no todos los sintetizadores incluyen este segundo paso y se limitan a crear una población.

Por lo tanto, para generar la población hay dos posibles opciones:

- Utilizar un sintetizador de población creado por terceros que genere una población a partir de una serie de datos iniciales específicos que difieren en cada programa. Para después obtener el *output* y modificarlo para que este se ajuste al simulador escogido y así poder trabajar con la población resultado.
- Generar de *motu proprio* la población utilizando y modificando alguno de los métodos que se proponen en los sintetizadores, o creando nuevos algoritmos para generar la población.

Se debe comentar la alta complejidad de la tarea, pues para crear una población se necesita de una gran cantidad de datos muy específicos que no se van a poseer y que no son fáciles de conseguir. Por ello, el primer paso consta de una investigación amplia de las fuentes de datos posibles para la región sobre la que se quiere simular. Es importante comentar que en función de la cantidad, calidad y tipología de los datos que se consigan reunir, se podrán realizar unas aproximaciones u otras. Las limitaciones de los datos serán determinantes para elegir entre unos programas u otros para generar la población sintética.

Dado que las empresas privadas no suelen compartir sus datos, las fuentes que ofrece el Estado en abierto, pueden ser un buen punto de partida en esta búsqueda de datos. Tanto a nivel de organismos oficiales, como pueden ser ministerios, institutos de estadística, ayuntamientos o sedes de los gobiernos regionales, que suelen ofrecer censos o cataistros, como a nivel de empresas públicas que también suelen ofrecer sus datos en abierto, como por ejemplo empresas públicas de transporte, de la vivienda, etc. También cabría mencionar como fuentes a muchas ONG o fundaciones que den acceso a sus datos.

Un tipo de fuente de datos clave que debería ser obligatoria, son las encuestas de movilidad (*travel survey*). De ellas se pueden extraer datos muy importantes relativos a los viajes diarios de los encuestados, pudiendo generar los planes de las personas sintéticas a partir de estos datos, así como ayudar a la generación de las familias. Las mejores fuentes de datos son aquellas que ofrecen sus datos en bruto sin procesar (*microdatos*) sobre las que poder realizar consultas para conseguir los datos específicos necesarios.

3.2.1. Sintetizadores de población

Dados un conjunto de datos demográficos pertenecientes a un subconjunto de la población real, un sintetizador de población busca ajustar esta muestra, seleccionando las variables más influyentes y ajustando sus pesos. Creando una población sintética a partir de las características de la población real que se han descubierto en los datos con los que se partía.

La síntesis poblacional puede ser considerada desde el punto matemático como un problema de optimización, que busca minimizar las diferencias entre las frecuencias generadas y las deseadas de las variables de control, buscando unos pesos óptimos para los hogares. Esto se puede formular mediante ecuaciones, mostrando distintas aproximaciones a la población sintética final.

Existen muchos enfoques distintos desde los que plantear la generación de población, pero todos se componen de dos pasos, un primer paso en el que se ajusta la población inicial obtenida de censos o encuestas de viajes a las limitaciones actuales y un segundo paso de selección de las casas y su posicionamiento en el mapa. Pese a esta parte común, la forma en que se llevan a cabo estos pasos difiere para cada sintetizadores, como reflejan Zhuge et al. en [44], los distintos enfoques para la sintetización poblacional se podrían agrupar en 5 grandes grupos:

- Modelos basados en la distribución demográfica. Este es el primer enfoque que se desarrolló para hallar poblaciones sintéticas y trata de compensar los problemas que suelen tener los censos, como que no estén completos o que éstos estén desactualizados, como es habitual en la mayoría de los países. Este enfoque busca ajustar los datos que surgen tras combinar los datos del censo con datos agregados actuales. Para llevarlo a cabo se buscan y seleccionan las variables más relevantes en la definición de una población, para después calcular su distribución marginal, con la que se halla la distribución demográfica de este subconjunto de la población. Asumiendo que la población total real tendrá las mismas proporciones que las del subconjunto, se crea una nueva población que siga una distribución y correlación similar, buscando también que parte de estos nuevos agentes creados tengan propiedades específicas coincidentes con los datos agregados. El resultado final es la creación de una población sintética en base a la distribución demográfica hallada en la pequeña muestra censal. Este enfoque pionero que hace uso del algoritmo *Iterative Proportional Fitting* (IPF) se le atribuye al trabajo de Beckman [45], y ha sido seguido por muchos otros investigadores con el afán de mejorarlo.
- Modelos basados en la simulación (*Simulation-based models*). Estos modelos en lugar de ajustar una única solución se centran en sintetizar poblaciones independientes, extrayendo agentes de la distribución conjunta de los atributos de la población real. Para ello usan, por ejemplo, el método de *Markov Chain Monte Carlo*, tal y como hicieron Farooq et al. [46] o como hicieron Felbermaire et al. [47]
- Modelos con un conjunto de datos restringido (*Data Constriction models*). Los modelos del primer enfoque pueden ser usados siempre que se tengan datos de un censo y datos agregados. Pero para los casos en los que no se dispone de estos datos censales surge este otro método, en el cual se parte solo de datos agregados y se trabaja mediante algoritmos iterativos semi-estocásticos que suplen los cruces entre los datos que faltan mediante procedimientos de prueba. Una aproximación por este método es el trabajo realizado por Gargiulo et al. [48].
- Modelos de optimización de la entropía (*Entropy Optimization Models*). Esta metodología puede formularse como un problema de optimización con restricciones, donde se intenta minimizar la diferencia entre las distribuciones de muestra ponderada y la distribución de la población conocida, esto se hace mediante una serie de variables de control a nivel de hogar y a nivel personal. Este método se basa por tanto en la estimación de los pesos de las encuestas de hogares, para que coincidan con las distribuciones de la población, buscando optimizar los pesos de las casas. Este método fue desarrollado por Bar-Gera et al. [49].
- Modelos basados en aptitud (*Fitness-based Models*). Este método puede partir de una lista inicial de casas vacía o con casas aleatorias, a partir de la cual se generará una lista de casas que se ajustan a distintos controles multinivel, mediante el añadido y el

eliminado de hogares durante las sucesivas iteraciones. Para ir evaluando el resultado y saber si se ha llegado al objetivo, se tienen unas tablas de recuento que cuantifican el número de hogares de cada tipo junto a unas tablas de control. Con este método se obvia determinar la distribución conjunta multidireccional necesaria en los métodos que hacen uso del algoritmo IPF, superando así los límites de los modelos que no permiten los controles multinivel. Este método es empleado en trabajos como el de Harland et al. [50] o el de Ma y Srinivasan [51].

Cada sintetizador tiene sus ventajas y sus desventajas y es el usuario el que debe valorar cuál de los sintetizadores disponibles se ajusta más a sus necesidades. Para tomar esta decisión conviene tener en cuenta los datos de los que se disponga, o el que la salida proporcionada sea de mayor o menor utilidad para su proyecto. También se puede considerar la complejidad de estos algoritmos, si la población que se desea generar es muy grande, el nivel de cómputo puede ser muy exigente y algunos sintetizadores no sean aptos para ello.

Por otra parte, también cabe comentar que hay poca información sobre los sintetizadores, de los pocos que se pueden encontrar referencias, no suelen tener su código en abierto.

A continuación, se muestran algunos ejemplos de sintetizadores de población:

- **Cemdap:** Este sintetizador generara una población basándose en modelos econométricos, es decir, haciendo uso de estadísticas en torno a fenómenos económicos para representar los comportamientos individuales de las personas. Para ello, necesita como *input* una serie de datos sobre los hogares, las personas, las infraestructuras y los datos geográficos, e introduce un nuevo método que combina las tablas de contingencia multidireccionales de forma genérica, para elaborar una población sintética y sus planes diarios. Este simulador se encuentra disponible en su propia web⁶ Chandra et al. “A Comprehensive Econometric Micro-simulator for Daily Activity-travel Patterns (CEMDAP)” [52]
- **PopulationSim:** Este sintetizador utiliza como base otro sintetizador de población llamado *PopSyn III*, al que añade distintas mejoras y subsana distintos errores. Partiendo de un pequeño número inicial de casas reales como ejemplo, hace uso de los procesos de balanceo de listas aplicados para diferentes niveles geográficos de *PopSyn III* y corrige ciertos errores de sus algoritmos. Genera como *output* una población sintética urbana y regional. El programa se encuentra disponible en su propia web⁷ [53] B. Paul et al. “PopulationSim”
- **ILUTE:** Este sintetizador hace uso del algoritmo IPF. Aporta un nuevo enfoque para manejar más atributos durante el proceso del IPF. Por otro lado, este simulador aporta una nueva solución para los casos en los que los datos de entrada no tengan establecida una relación entre las personas y las casa, creando mediante su algoritmo esta relación.
- **PopGen:** Este sintetizador aportó un nuevo algoritmo heurístico para generar iterativamente una población sintética representativa, que busca optimizar algoritmos de otros sintetizadores anteriores que únicamente se centraban en ajustar las distribuciones de los atributos a nivel de hogar. Para hacer esto PopGen aporta una

⁶ Página web del sintetizador de población [Cemdap](#)

⁷ Página web del sintetizador de población [PopulationSim](#)

técnica novedosa que ajusta los marginales de las variables de los hogares y de las personas simultáneamente. El software esta disponible en su página web⁸ [54]

3.2.2. Nuevo enfoque en la sintetización de población; Estructuras sociales

Al revisar los distintos enfoques de la generación de poblaciones sintéticas, se observó que la mayoría de ellos se centran en estudiar mediante algoritmos los factores que definen la muestra del censo de la que parten, para luego extrapolar estos conocimientos a la población total, dejando de lado las estructuras sociales de la población. En el enfoque de este trabajo, aunque sigue partiendo de datos muestrales de un censo, se decidió seguir una aproximación distinta, centrándose en un análisis de las estructuras sociales, dejadas de lado por el resto, para posteriormente generar la población en base a estas estructuras.

Aunque es verdad que las estructuras familiares difieren dependiendo de la región en la que se realice el estudio, hay estructuras básicas como, por ejemplo, la familia compuesta por unos padres y sus hijos, que están presentes en la mayoría de las culturas, en distintos porcentajes de aparición, pero están presentes. Por otro lado, la reducción de las familias a la raíz y la creación de nuevas familias a partir de agregación de nuevos individuos a estas raíces, simplificaría la creación de las familias, pudiendo adaptarse mejor los resultados a otras culturas y regiones. Aunque para la aplicación de este modelo a otras regiones, será imprescindible contextualizar las estructuras identificadas con las estructuras sociales características de la nueva región, las cuales podrían incluir nuevas tipologías que se deberán incorporar al algoritmo. Por lo tanto, para desarrollar este enfoque se inició este proyecto con un primer estudio social, con el que se pretenden identificar las estructuras sociales base, para posteriormente escalar el proyecto a otras regiones.

Una ventaja que se aporta en este proyecto surge del uso de algunos parámetros más generales extraídos del Instituto Nacional de Estadística (INE), pues estos no provienen de una muestra de la población real, si no que provienen de la población real total. Con esto se reduce el riesgo a obtener una población sesgada, ya que cuando solo se parte de datos tomados de una muestra, en el caso de que esta muestra no refleje correctamente la diversidad y complejidad de la población total, podría obtenerse una población sesgada inválida, como puede pasar en otros sintetizadores que basan completamente su modelo en la muestra.

Este trabajo consta de un estudio de las tipologías de familia existentes previo a la creación del software. Con este estudio se pretenden identificar los distintos tipos de familias y su porcentaje de aparición para la región en estudio, que en este caso se trata de Leganés, ciudad en la que se sitúa uno de los campus de la Universidad Carlos III de Madrid.

Debido a que este sintetizador genera los planes diarios de las personas, va a ser necesario recopilar datos sobre el comportamiento de esta población, la cantidad de desplazamientos diarios, transportes que se utilizan, horarios fijos que sigan semanalmente, distancias recorridas y un largo etc. Toda esta información se puede obtener mediante encuestas de movilidad también conocidas como encuestas de viaje (*travel survey*). Las encuestas de viaje analizan el comportamiento de las personas, recopilando información sobre el hogar, sobre los viajes realizados en un día determinado y sobre los propios encuestados. También se pueden obtener datos de matrices de comutación origen/destino que, como su nombre indica, ofrecen información por pares de origen y destino de los flujos de personas, viajes clasificados por el

⁸ Página web del sintetizador de población [PopGen](#)

propósito, como por ejemplo, ir al trabajo. Gracias a estas matrices se puede observar los tipos de viajes que se producen entre las distintas zonas de la región que se estudia.

Por último, va a ser necesario obtener datos sobre los tipos de servicios que se ofertan en la zona geográfica del estudio, es decir, explicitar el uso de los edificios pudiendo ser de uso residencial, de uso comercial, industrial, etc. Esto permitirá conocer a qué lugares suelen dirigirse las personas dentro del mapa cuando quieren realizar un tipo de actividad concreto.

En caso de no conseguir todos los datos necesarios, aunque no es lo deseable, se pueden intentar fabricar algunos de ellos. Se podrían, por ejemplo, utilizar distribuciones aleatorias en una zona determinada, para generar los datos necesarios que permitan simular.

3.2.3. Descripción del programa: *The Synthetic Population*

El sintetizador de población aquí desarrollado está dirigido específicamente para ser usado en Matsim, ya que el fichero que se obtiene como salida está estructurado y contiene los datos necesarios que utiliza Matsim para llevar a cabo una simulación. Sin embargo, aunque el *output* este dirigido a Matsim, tanto la población como los planes que se generan son válidos para cualquier otro simulador basado en agentes y por ello, si se modifica el *output* restructurando los datos, estas familias y planes generados se podrían utilizar en cualquier otro simulador basado en agentes.

The Synthetic Population está compuesto por dos módulos principales, el primero nombrado el Familiador, genera familias sintéticas a partir de los datos extraídos del censo de Leganés, del catastro y del INE. El segundo módulo denominado el Planificador, utiliza los datos extraídos de una encuesta de movilidad realizada por el Consorcio Regional de Transportes de Madrid (CRTM), para asignar planes de familias reales encuestadas a familias sintéticas generadas por el Familiador. En la sección 4. Análisis correspondiente al análisis del sistema, se expondrán los casos de uso y los requisitos funcionales y no funcionales del programa.

En la Figura 15, se pueden apreciar los dos módulos mencionados de *The Synthetic Population*, el primer módulo toma datos de INE para obtener las estructuras familiares y sus porcentajes de aparición, toma datos del censo de Leganés para obtener la población exacta del municipio y, por último, toma datos del catastro para poder asignar casas reales a las personas creadas. Con todos estos datos de partida, el Familiador genera familias sintéticas hasta acabar con el censo de Leganés, obteniendo como resultado final la población sintética del municipio.

Por otro lado, el Planificador toma la población sintética sin planificación diaria asignada del Familiador para, a continuación, asignar a cada familia sintética una familia real similar tomada de la encuesta de viajes, la cual posee planes diarios. Una vez seleccionadas las familias, el programa transfiere los esqueletos de los planes de las familias reales a las familias sintéticas, esqueletos que no contienen las ubicaciones de los destinos, pues estos se calcularán mediante un método que se explica más adelante. Una vez halladas las coordenadas destino de cada plan, estas coordenadas se agregarán a los esqueletos completando los planes. Este proceso se realizará para todas las familias sintéticas, obteniendo finalmente como *output* un archivo XML con la población sintética terminada, que estará lista para ser utilizado por MATSIM para realizar la simulación.

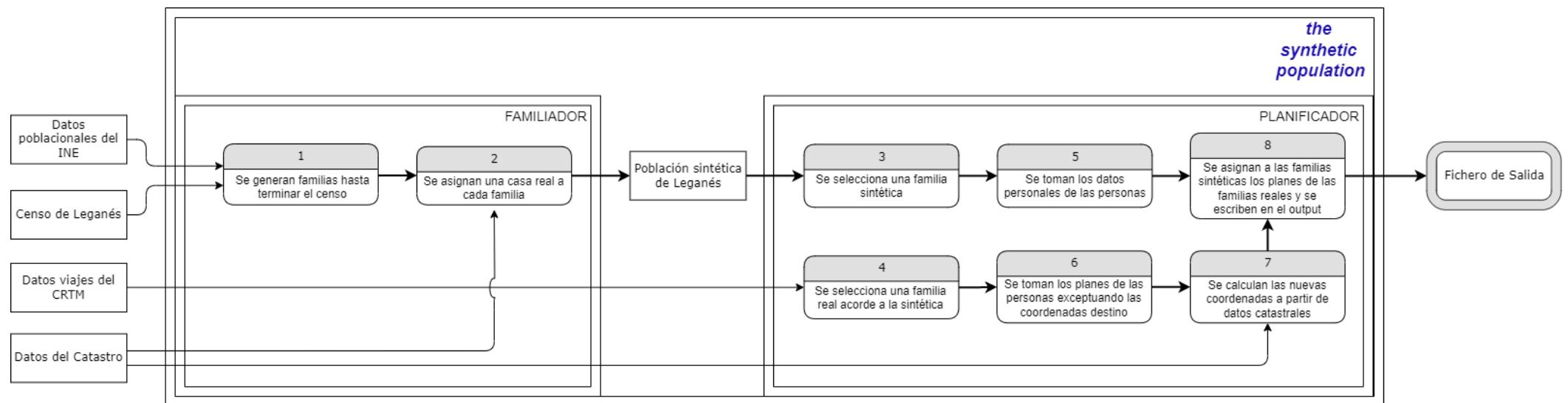


Figura 15. Diagrama The Synthetic Population

El manual de usuario donde se enumeran todos los datos específicos necesarios para el correcto funcionamiento del software, así como de las estructuras de los archivos que se tienen que incluir con los datos del censo, del catastro y de la encuesta de movilidad, se encuentran en el Anexo B, que abarca ampliamente todas estas cuestiones.

3.2.4. Datos y Fuentes

Este apartado contiene las fuentes de datos utilizadas para este proyecto, de las que se detallan los procesos de obtención y de preprocesado, así como de otras fuentes de datos iniciales que finalmente no llegaron a ser utilizadas. Este punto es ampliamente desarrollado en el Anexo A, en el cual se detallan los procesos de adquisición de datos.

3.2.4.1. Censo de Leganés

La página del Ayuntamiento⁹ de Leganés contiene un apartado de estadísticas donde se encuentran los datos del censo, estos datos están recogidos por barrios o por distritos. Debido a una serie de problemáticas que se explican en el Anexo A, finalmente se decide utilizar el censo por distritos.

El preprocesado del censo diferirá según la estructura elegida por cada ayuntamiento, pues la presentación de estos datos no está normalizada, aun así, la dificultad no es alta. El resultado de este preprocesado debe conseguir los datos relativos al número de personas para cada edad, divididas por sexo y ordenadas de menor a mayor edad.

3.2.4.2. Datos estadísticos poblacionales del INE

Aunque inicialmente se encontraron multitud de fuentes de datos posibles, como las aportadas por algunos ayuntamientos, por la Comunidad de Madrid, e incluso por páginas web que recopilaban datos de distintas fuentes. Finalmente, se decide utilizar casi de forma exclusiva los datos del Instituto Nacional de Estadística (INE), debido a la cantidad y calidad de los datos que aportaba, pese a la falta de datos específicos del municipio de Leganés, que no estaban disponibles por temas de protección de datos.

El INE es una un organismo que recoge una gran cantidad de datos de distinta índole, por ello, para obtener datos específicos hay que identificar que encuestas contienen los datos requeridos. La fuente principal de datos para este proyecto fue la encuesta continua de hogares (ECH), una investigación que ofrece información anual sobre las características demográficas básicas de la población y de los hogares que componen, mientras que la encuesta de población activa (EPA) fue una fuente secundaria, la EPA es una investigación continua realizada trimestralmente dirigida a las familias, cuya finalidad principal es obtener datos de ocupados, parados e inactivos.

Dentro de cada encuesta, el INE proporciona una serie de tablas con unas consultas prefijadas. Por otra parte, en caso de necesitar datos más específicos, el INE también ofrece unos microdatos que se pueden descargar, para que los usuarios, mediante consultas, puedan extraer datos específicos necesarios para sus proyectos. Estos datos en el caso de la ECH contienen una pequeña muestra representativa de la población española e incluye una gran cantidad de atributos por los que filtrar para obtener la población objetivo.

⁹ Página Web del ayuntamiento de [Leganés](#)

Para las personas que no tienen conocimientos de programación, el INE tiene otra herramienta para poder obtener datos específicos que las tablas prefijadas no contienen. Esta herramienta llamada *Generador de Tablas a Medida*, permite filtrar los datos paso a paso con una serie de filtros como el ámbito geográfico, el colectivo sobre el que se realiza la consulta y los propios datos que se desean, consiguiendo finalmente una tabla definida por el usuario, con los datos requeridos.

A través de distintas consultas, utilizando el lenguaje de programación R y minoritariamente el *Generador de Tablas a Medida*, se obtuvieron los datos necesarios para representar las tipologías de hogar identificadas en el estudio sociológico que se desarrolló en este trabajo.

3.2.4.3. Datos de los inmuebles del Catastro

Matsim necesita las coordenadas exactas de los destinos o en su defecto el nodo de la red viaria más próximo para calcular los desplazamientos de los agentes. Por ello, era necesario obtener las coordenadas exactas de los edificios de Leganés. Aunque inicialmente se pensó en utilizar los datos de las infraestructuras que proporciona *OpenStreetMap*, al encontrar que el Catastro contenía los datos oficiales del Estado, datos más rigurosos que los facilitados por OSM, se decide tomar al Catastro como fuente de datos principal.

Los bienes inmuebles del catastro están clasificados según el uso que se les da, en la Figura 16. se muestran todas las categorías identificadas por el Catastro. Dado que los datos del catastro

eran necesarios en dos ocasiones, una primera vez para obtener las viviendas de las familias en el módulo del Familiador y en una segunda ocasión para obtener las coordenadas de los destinos de los planes en el Planificador. Estos datos se filtraron por distintos usos en cada parte del programa, dejando los datos de los inmuebles necesarios en cada ocasión.

Una vez se obtuvieron los edificios necesarios, ya solo restaba borrar las columnas de datos innecesarios, quedándose solo con las pertinentes. Los datos necesarios en ambas consultas eran prácticamente idénticos a excepción de una variable.

CODIGO	USO
A	Almacén -Estacionamiento
V	Residencial
I	Industrial
O	Oficinas
C	Comercial
K	Deportivo
T	Espectáculos
G	Ocio y Hostelería
Y	Sanidad y Beneficencia
E	Cultural
R	Religioso
M	Obras de urbanización y jardinería, suelos sin edificar
P	Edificio singular
B	Almacén agrario
J	Industrial agrario
Z	Agrario

Figura 16. Tabla de los usos de los bienes inmuebles [55]

Para este trabajo se han utilizado las siguientes columnas:

- Código de la parcela identificativo
- Distrito municipal en el que se sitúa el inmueble (Utilizado solo para las viviendas)
- Metros cuadrados del inmueble
- Coordenada x en coordenadas UTM zona 30N (EPSG:23030)
- Coordenada y en coordenadas UTM zona 30N (EPSG:23030)

3.2.4.4. Datos de los desplazamientos de la encuesta de movilidad del CRTM

Para la creación de la planificación diaria de los hogares sintéticos, se utilizaron como base los datos de la encuesta de movilidad del Consorcio Regional de Trasportes de Madrid (CRTM) [55], que estaban disponibles en su página web¹⁰.

El preprocesado de estos datos consistió en la división de los hogares en ocho archivos, establecidos en base a una serie de parámetros que definían distintos tipos de movilidad, para poder así asignar planes reales a las familias sintéticas. Por otra parte, este preprocesado también incluye la modificación de algunos viajes que presentaban incoherencias, ya que los datos descargados de consorcio de trasportes no estaban completos.

Muchas personas contaban con solo un viaje diario, para arreglar esto, a todas estas personas se les agrego un viaje de vuelta a su vivienda y que así, tuvieran como mínimo dos viajes diarios. Por otra parte, a algunas personas le faltaban planes intermedios que hicieran de nexo entre dos planes. Para corregir esto, se modificaron los orígenes de los segundos planes, para que coincidieran con el destino del primer plan, en la Tabla 1 se muestra un ejemplo de esto.

Cambio de orígenes para mantener la coherencia				
Acción	Origen 1	Destino 1	Origen 2	Destino 2
-	Domicilio	Trabajo	Supermercado	Domicilio
Cambio origen 2	Domicilio	Trabajo	Trabajo	Domicilio

Tabla 1. Cambio de orígenes

Algunas familias estaban compuestas únicamente por niños, por ello estas familias y sus planes fueron eliminadas de los datos. Por otro parte, muchas familias incluían personas sin planes que se pasaban el día entero en su domicilio, como estas personas eran necesarias para entender la composición de las familias, no fueron eliminadas.

Dado que únicamente se habían encuestado a 2407 personas del municipio de Leganés, para obtener una base de datos mayor sobre la que iniciar la creación de planes, se agregaron a las personas encuestadas de municipios aledaños a Leganés que compartían muchas similitudes con este, como la renta, la ubicación con respecto a Madrid y la menor cantidad de empleos, que hace que parte de su población salga de estas ciudades para ir a trabajar los días de diario, las ciudades escogidas fueron Getafe, Alcorcón y Móstoles. Esto implicaba tener que modificar los orígenes y destinos de la planificación diaria de estas personas, para ello, en los desplazamientos realizados dentro del municipio se cambiaba las antiguas poblaciones, por la de Leganés, en el caso de viajes a fuera del municipio, se modificaba la ciudad de residencia por Leganés, manteniendo la otra ciudad a la que se desplazaban, a excepción de los viajes que se daban entre las antiguas ciudades de residencia y Leganés. En estos casos, para evitar que, al cambiar la ciudad de origen por Leganés, se dieran viajes de Leganés a Leganés, cambiando el tipo de viaje de extramunicipal a intramunicipal, se intercambiaron los municipios de origen y destino.

Para facilitar la comprensión de los tres casos a modificar relatados, se incluye en la Tabla 1 un ejemplo de cada uno de ellos.

¹⁰ Apartado de la página web del [CRTM](#) con los datos de la encuesta

Cambio de municipio para encuestados de otras poblaciones						
Persona	Acción	Ciudad en la que vive	Origen 1	Destino 1	Origen 2	Destino 2
1	-	Móstoles	Móstoles	Móstoles	Móstoles	Móstoles
	Cambio de Municipio	Leganés	Leganés	Leganés	Leganés	Leganés
2	-	Móstoles	Móstoles	Madrid	Madrid	Móstoles
	Cambio de Municipio	Leganés	Leganés	Madrid	Madrid	Leganés
3	-	Móstoles	Móstoles	Leganés	Leganés	Móstoles
	Intercambio municipios	Leganés	Leganés	Móstoles	Móstoles	Leganés

Tabla 2. Cambio municipio

Cabe señalar que todo el preprocesado de los datos de la encuesta de movilidad nombrados en este apartado fueron realizados por Álvaro Delgado y son ampliamente desarrollados con una mayor precisión en su TFG [56]

3.2.5. Creación de las familias: El Familiador

El primer paso del trabajo consistió en generar la población, es decir, generar los futuros agentes que se desplazarán durante la simulación. Sin embargo, antes de empezar con la programación, hay que conocer a las personas que hay que generar para entender cómo se desplazan. La forma en que la gente se mueve se puede explicar mediante una serie de factores que definen sus comportamientos y que hacen que la movilidad difiera entre unas personas y otras.

Los humanos son seres sociales en constante interacción con otros seres humanos. Estas interacciones se orientan por una serie de reglas culturales que define la sociedad. Aunque estas reglas difieran según la cultura propia de cada contexto, todas comparten que la movilidad de las personas está influenciada por su núcleo social cercano. Las personas planifican su día a día en torno a unos objetivos básicos sobre los que se incluyen a las personas que les rodean.

Algunas de las reglas que definen la sociedad hacen referencia a los modelos de agrupamiento. Amigos o compañeros de trabajo van a influir en la movilidad, sin embargo, la mayor influencia la ejerce la familia, y cuando se utiliza el término familia, en realidad este se refiere al hogar, a las personas con las que se convive. En la mayoría de los casos esta convivencia se realiza con personas con las que se comparten lazos de sangre, pero también existen otras tipologías de hogar en los que no se comparten lazos o incluso tipologías de hogares en los que conviven varias familias dentro de un mismo hogar. Es cierto, que, para los dos últimos tipos de hogar, las interconexiones en los planes de los miembros del hogar pueden ser menores que los de una familia, pero, aun así, pese a no compartir lazos de sangre, la convivencia hace que se formen

dinámicas familiares, como por ejemplo salir a tomar algo, o hacer la compra. Punto por el cual, tiene sentido llamar a estos hogares familias y, por tanto, tratarlas como tales, aunque se gestionen de modo diferente.

Como la asignación de planes se va a realizar por hogares, la generación de la población sintética no se va a realizar persona a persona, sino que se irán creando hogares, intentando tener en cuenta las múltiples tipologías existentes.

Aunque las personas con las que se convive son las que mayor influencia ejerce sobre la planificación de las actividades, como se ha comentado anteriormente, otros familiares, amigos, conocidos e incluso compañeros de trabajo también van a influir, como, por ejemplo, el compartir coche con un compañero de la oficina para desplazarse acompañado al lugar de trabajo. Sin embargo, determinar cómo influyen estas relaciones en la movilidad es mucho más complejo de representar y por ello en este primer prototipo no se han llegado a implementar, con la idea de trabajar en ello y profundizar más en un segundo prototipo.

3.2.5.1. Estudio sociológico de los hogares

El hogar es el eje generador desde el que se está trabajando y por extensión las personas que conviven en él. Dado que la movilidad viene en parte definida por el hogar, es necesario realizar un análisis de las personas para poder recrear sus patrones de movilidad.

Este estudio tiene que analizar cómo son estos hogares, qué los diferencia, qué tipos de desplazamientos realizan, centrándose por tanto en analizar un conjunto de factores escogidos por una serie de argumentos que justifican su relevancia para la composición y la movilidad de las familias.

- **La edad:**

» **Composición:** Las personas van pasando por distintas etapas de la vida, la composición del hogar difiere en cada una de estas etapas. La niñez se caracteriza por estar basada en un vínculo de dependencia de los hijos hacia los padres y una planificación de los desplazamientos impuesta por la sociedad (colegio) o los adultos (ocio, periodos de vacación, etc.). Durante la adolescencia y juventud las personas cobran autonomía en sus vidas y por ende de sus desplazamientos. En esta etapa es donde se generan la mayoría de las relaciones amistosas y amorosas, que se suelen dar con otras personas de edad similar y que acabarán desembocando en la constitución de los grupos de amigos y futuras familias, lo que hace que las relaciones se sigan dando con personas de edades similares. En la madurez aparece una necesidad personal y una presión social para emanciparse y residir en otro domicilio lo que supone “cambiar de familia” según la concepción de este trabajo. Por todo ello, la edad es un factor muy importante para diferenciar unos hogares de otros.

» **Movilidad:** Por otro lado, la movilidad de las personas también cambia radicalmente según la etapa de la vida, la vitalidad de la gente joven, la mayor cantidad de tiempo libre y la falta de responsabilidades hace que los jóvenes tengan una mayor movilidad. Inicialmente, se desarrolla en el entorno cercano al hogar, pero progresivamente se va expandiendo, realizándose planes cada vez más lejos de casa. En la adultez, el aumento de las responsabilidades reduce el tiempo libre,

se estandarizan muchos desplazamientos, cambiando la forma en que se desplazaban estas personas durante su juventud. Finalmente, en la vejez, tras la jubilación se tiene mucho más tiempo libre, sin embargo, la vitalidad disminuye lo que limita la cantidad de planes que se hacen. Cada etapa de la vida tiene una movilidad diferente, que se deberá estudiar.

- **El nivel de estudios:** Un mayor nivel de estudios tiene una correlación positiva con el factor renta per cápita y, por tanto, parte de las razones para su inclusión serán analizados en el factor de la renta per cápita.

» **Composición:** Un mayor nivel de estudios, implica una dilación en la edad de emancipación de los hijos. Si se han cursado estudios superiores, el acceso al mundo laboral se retrasa y con ello, probablemente se posponga la formación de la familia, una familia en la que la pareja suele tener el mismo nivel académico.

El mejor predictor de continuidad en los estudios de los hijos, es el nivel alcanzado por los padres. Por tanto, aquellas familias que tienen estudios superiores previsiblemente tendrán hijos que permanecerán más tiempo en el hogar, realizando desplazamientos a sus centros de estudio. La tendencia natural es que la independencia de los hijos coincida con la finalización de sus estudios y el logro de un empleo.

» **Movilidad:** La movilidad de los estudiantes es previsible y regular durante los períodos activos del calendario docente. En el caso de zonas rurales o ciudades pequeñas, acceder a los estudios superiores suele suponer un traslado de los estudiantes a nuevas ciudades, buscando domicilios cercanos a sus nuevos centros educativos. Una vez instalados en el nuevo domicilio, si se exceptúan los frecuentes viajes a sus ciudades de origen para ver a familia y amigos, los trasladados volverán a ser regulares y previsibles.

Las personas con distintos niveles de estudios tienden a tener gustos culturales y de consumo acordes con los diferentes niveles, variando los desplazamientos para este tipo de actividades según sus preferencias.

- **Situación laboral de los integrantes:** Estadísticamente, una mayor cualificación o nivel de estudios correlaciona con un índice de desempleo menor y unos ingresos superiores. La falta de estabilidad laboral de la última década hace más impredecible la continuidad de los proyectos personales de los integrantes de las familias.

» **Composición:** Aunque hay personas que estudian y trabajan, es más usual terminar los estudios antes de acceder al mundo laboral. Para independizarse, es necesario tener unos mínimos ingresos y ahorrar, optando a compartir gastos con amigos o con una pareja. La decisión de tener hijos viene fuertemente condicionada por la estabilidad económica.

» **Movilidad:** La actividad laboral es una fuente diaria de desplazamientos en los días laborales. Pudiendo distinguir trabajos con una movilidad regular, que realizan un desplazamiento al lugar de trabajo que no varía y trabajos con constantes movimiento dentro de la ciudad e incluso fuera de esta, en donde las rutas varían día a día. La crisis sanitaria de 2020 supuso un aumento significativo del teletrabajo, pasados los picos de máximas restricciones, este teletrabajo se ha mantenido en

muchos de los trabajos que poseían una movilidad regular a la oficina, reduciendo drásticamente los desplazamientos.

Dado que las personas con empleo tienen unos ingresos fijos, estos pueden permitirse gastar dinero los días no laborables y en los períodos vacacionales, aumentando su tasa de desplazamientos y la heterogeneidad de éstos.

La gente que no tiene un empleo se divide en dos grupos, un primer grupo de personas que no están en activo, que suelen ser personas mayores y que tiene una movilidad particular y, un segundo grupo de personas desempleadas que buscan trabajo y realizan algunos desplazamientos para su búsqueda, pero que no tienen ni la cantidad ni la regularidad de viajes de las personas empleadas.

- **La nacionalidad/cultura:** Hay una correlación entre la educación, la renta de los hogares y la nacionalidad. Esta correlación difiere entre países desarrollados y países en vías de desarrollo. Mientras que las personas que vienen de los países desarrollados suelen emigrar al encontrar un trabajo cualificado gracias a su nivel de estudios, esto les garantiza unos altos ingresos en el nuevo país y una movilidad asociada acorde, aunque las políticas migratorias no siempre facilitan esto. Por otro lado, las personas que emigran de países en vías de desarrollo vienen buscando unos ingresos que no pueden obtener en su país de origen, estos migrantes no suelen tener estudios superiores y por tanto no tendrán acceso a trabajos que requieran una alta cualificación, lo que suponen una movilidad claramente diferenciada entre migrantes.

➤ **Composición:** El origen cultural de las personas tiene un gran efecto diferenciador en las familias. Puede generar cambios en la estructura de las familias, en el número de hijos deseables, en algunos casos admitiendo la poligamia o favoreciendo la convivencia en un mismo espacio de una familia extensa (por ejemplo, abuelos, hijos y sus consortes, nietos junto con los hijos de estas parejas). También afecta al modo y momento de la emancipación de los hijos. Un ejemplo de ello son las tradiciones que algunas familias de cultura musulmanas o de etnia gitana siguen practicando.

➤ **Movilidad:** La religión y el número de desplazamientos a sus respectivos centros de culto, las festividades y celebraciones, los hábitos de consumo y desplazamiento para su adquisición, la concepción del ocio y muchos más aspectos pueden afectar de manera diferencial a la movilidad de personas pertenecientes a distintas culturas o etnias. Esto dependerá del grado de práctica de tradiciones o rituales asociados a la cultura de origen de las familias.

- **La renta per cápita:**

➤ **Composición:** La composición de las familias difiere entre países desarrollados y países en vías de desarrollo, en los países más pobres se suelen tener muchos más hijos, dado que estos sirven como fuente de ingresos al ponerse a trabajar a una edad temprana, también que la mortalidad infantil sea más alta hace que se tengan más hijos para que alguno llegue a la edad adulta. La situación española no tiene las desigualdades de los países menos desarrollados, pero, aun así, las familias con menos ingresos suelen tener más hijos que las de mayores rentas. La clase media no puede permitirse muchos hijos si desean mantener su nivel económico y no

renunciar al nivel de vida. La crianza de los hijos en la clase media es mucho más prolongada pues se espera en muchos casos que estos lleguen hasta la universidad, también es costosa ya que se destinan en ellos una parte importante de los recursos económicos de la familia para su formación, lo que limita el número de hijos a los que se puede dar una educación satisfactoria. Por otro lado, familias adineradas que si pueden permitirse tener más hijos tampoco suelen tenerlos.

» **Movilidad:** La movilidad de las familias varía según los ingresos de los que dispongan, mayores rentas se suelen traducir en un número mayor de desplazamientos para actividades de trabajo (a otros lugares e incluso a otros países), educativas (adquisición de idiomas, experiencias de autonomía, etc.) o de ocio (viajes y vacaciones, restaurantes, espectáculos, etc.) especialmente en los fines de semana, donde se pueden ver desplazamientos que cubren mayores distancias.

- **El tamaño de las viviendas:** A no ser que medie una herencia de por medio, el tamaño de las viviendas tiene una correlación más que significativa con la renta per cápita.

» **Composición:** Existe una relación directa entre el tamaño de las casas/número de habitaciones con el tamaño de las familias, aunque el precio del metro cuadrado del barrio en el que se ubica la casa hará que esta relación varíe. Un barrio formado en su mayoría por chalets no contendrá un gran número de familias unipersonales.

» **Movilidad:** Aunque en todos los barrios hay casas de tamaños diversos, algunos barrios concentran más casas de un cierto tamaño que otros, soliendo ser los barrios con menores ingresos, los más céntricos y los más antiguos, los que contienen una mayor cantidad de viviendas pequeñas, lo que hace que la densidad de población sea mucho mayor. Al tener una mayor densidad de población, estos barrios tendrán una movilidad distinta en la que los transportes colectivos estarán más presentes, esta situación se invierte en los barrios que tiene una menor densidad de población.

No todos los patrones mencionados pudieron ser finalmente utilizados, pues obtener datos de alguno de ellos era muy complejo y se decidió dejar su implementación para un segundo prototipo.

3.2.5.2. Tipologías de hogares

A partir del análisis de los factores mencionados se estudió la composición de los hogares, distinguiendo una serie de tipologías que debían ser implementadas. No obstante, para poder introducir estas tipologías, era necesario saber cuál era su grado de aparición. Gracias a una tabla del INE, la cual se muestra en la Figura 17, se obtuvieron los porcentajes de una serie de tipologías de hogar más generales que estaban divididas por el número de habitantes. Por lo que, para estructurar y organizar la generación de los hogares, se utilizó la división por número de convivientes mostrada en la tabla, para luego una vez elegido el número de personas que vive en la casa, elegir una de las tipologías disponibles para ese número de convivientes.

	Total	1 persona	2 personas	3 personas	4 personas	5 personas	6 personas	7 personas	8 personas o más
2019									
Madrid, Comunidad de									
Total (tipo de hogar)	2.612,6%	672,2%	789,3%	525,3%	457,8%	102,4%	40,5%	11,1%	14,0%
Hogar unipersonal	672,2%	672,2%
Hogar monoparental	262,2%	..	173,9%	72,0%	14,8%	1,5%
Pareja sin hijos que convivan en el hogar	531,5%	..	531,5%
Pareja con hijos que convivan en el hogar. Total	854,5%	388,7%	387,3%	65,2%	11,1%	0,7%	1,6%
Pareja con hijos que convivan en el hogar. 1 hijo	388,7%	388,7%
Pareja con hijos que convivan en el hogar. 2 hijos	387,3%	387,3%
Pareja con hijos que convivan en el hogar. 3 o más hijos	78,5%	65,2%	11,1%	0,7%	1,6%
Núcleo familiar con otras personas que no forman núcleo familiar	120,7%	46,8%	36,7%	20,5%	12,7%	1,3%	2,7%
Personas que no forman ningún núcleo familiar entre sí	106,3%	..	83,9%	17,8%	3,9%	0,7%
Dos o más núcleos familiares	65,2%	15,1%	14,5%	16,7%	9,1%	9,7%

Figura 17. Tabla hogares por tipo y número de personas [57]

Partiendo de los tipos de familias más generales que aparecen en la Figura 17, a través de un profundo estudio sociológico se identificaron nuevas tipologías que surgen al subdividir estos tipos más generales en distintas subclases más específicas. Con lo que se busca conseguir una tipología más detallada y realista que no agrupe distintos tipos de familias en una sola clase.

Entre los tipos de hogar más generales utilizados por el INE, aparece el término núcleo familiar. Éste hace referencia a una concepción de la familia limitada a los vínculos de parentesco más cercanos y es utilizado ampliamente en sociología. Los núcleos familiares son, por tanto, familias que están compuestas por:

- Una pareja o matrimonio que convive sin hijos.
- Una pareja o matrimonio que convive con hijos.
- Una madre que convive con hijos.
- Un padre que convive con hijos.

Estos núcleos familiares pueden vivir solos o junto a otras personas ajenas al núcleo familiar en el mismo hogar, e incluso se pueden dar hogares compuestos por varios núcleos. Por ello, para la generación de hogares más complejos, que no pueden ser fácilmente identificado con la subdivisión de los tipos de hogares generales, se pueden utilizar estos núcleos familiares como raíz a la que se agregan nuevos individuos.

Definidas ya las tipologías, se estableció la estructura final para el *Familiador*. El programa creado inicia la generación de los hogares con un número aleatorio, que se utiliza para elegir el número de personas que componen la familia en función de sus probabilidades, mostradas en la Tabla 3. Cabe destacar que se han unido las familias de 7 con las familias de 8 o más, dado que los porcentajes de aparición eran muy bajos. A continuación, escogido ya el número de personas de las que se compone el domicilio, el *Familiador* elige entre las tipologías de hogar existentes para ese número de personas.

Personas	1	2	3	4	5	6	7 o más
Porcentaje	25.7%	30.2%	20.1%	17.6%	3.9%	1.5%	1%

Tabla 3. Porcentajes de hogar por número de personas

El resultado final tras el estudio sociológico de las tipologías generales, con el que se identificaron los tipos de hogares a implementar, queda reflejado en el diagrama de la Figura 18 y en la Figura 19, donde aparecen todas las tipologías que se llegaron a identificar.

El estudio de las tipologías mostrado consiguió identificar una gran cantidad de tipologías de familias. En sucesivos proyectos se espera continuar el estudio de los microdatos del INE, para obtener las estadísticas necesarias que permitieran identificar todas las tipologías halladas en porcentajes significativos. Esto no es una tarea fácil, por poner un ejemplo, para incluir la clase piso compartido, habría que estudiar que subclases se pueden encontrar, ¿Tienen edades parejas los integrantes del hogar? ¿Comparten algún vínculo varios integrantes de la casa? ¿Qué género tienen los integrantes? ¿Es un piso mixto o un piso de *amigos*?

El estudio de las familias compuestas por 1 o 2 personas se completó quedando pendiente continuar el estudio para el resto de las familias. La significatividad de las tipología descritas en los diagramas deberían ser verificadas mediante consultas a los datos del INE. Consultas con las que se podrían, a su vez, identificar la existencia de tipologías que hubiera sido pasada por alto.

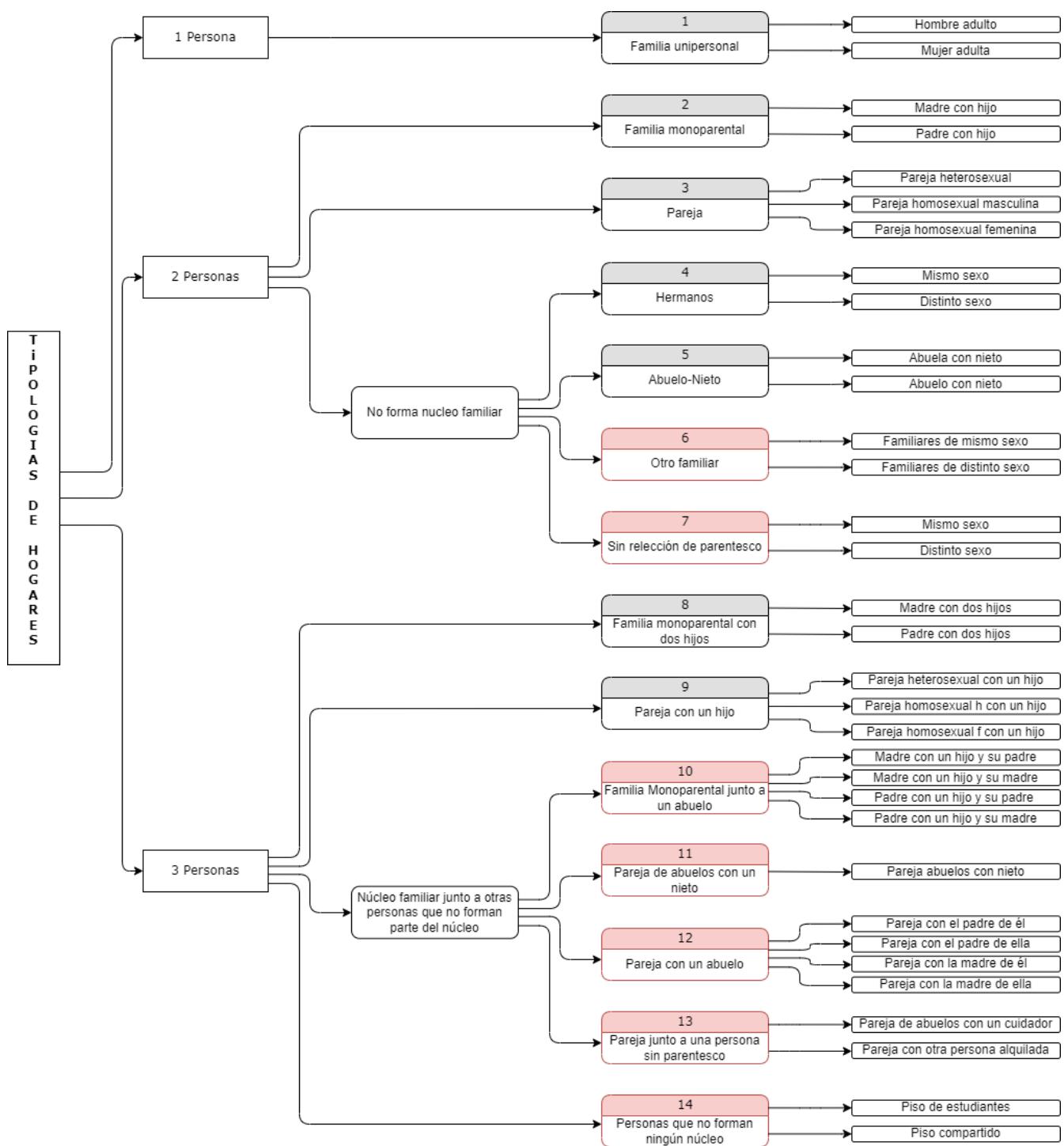


Figura 18. Diagrama tipologías de hogar 1

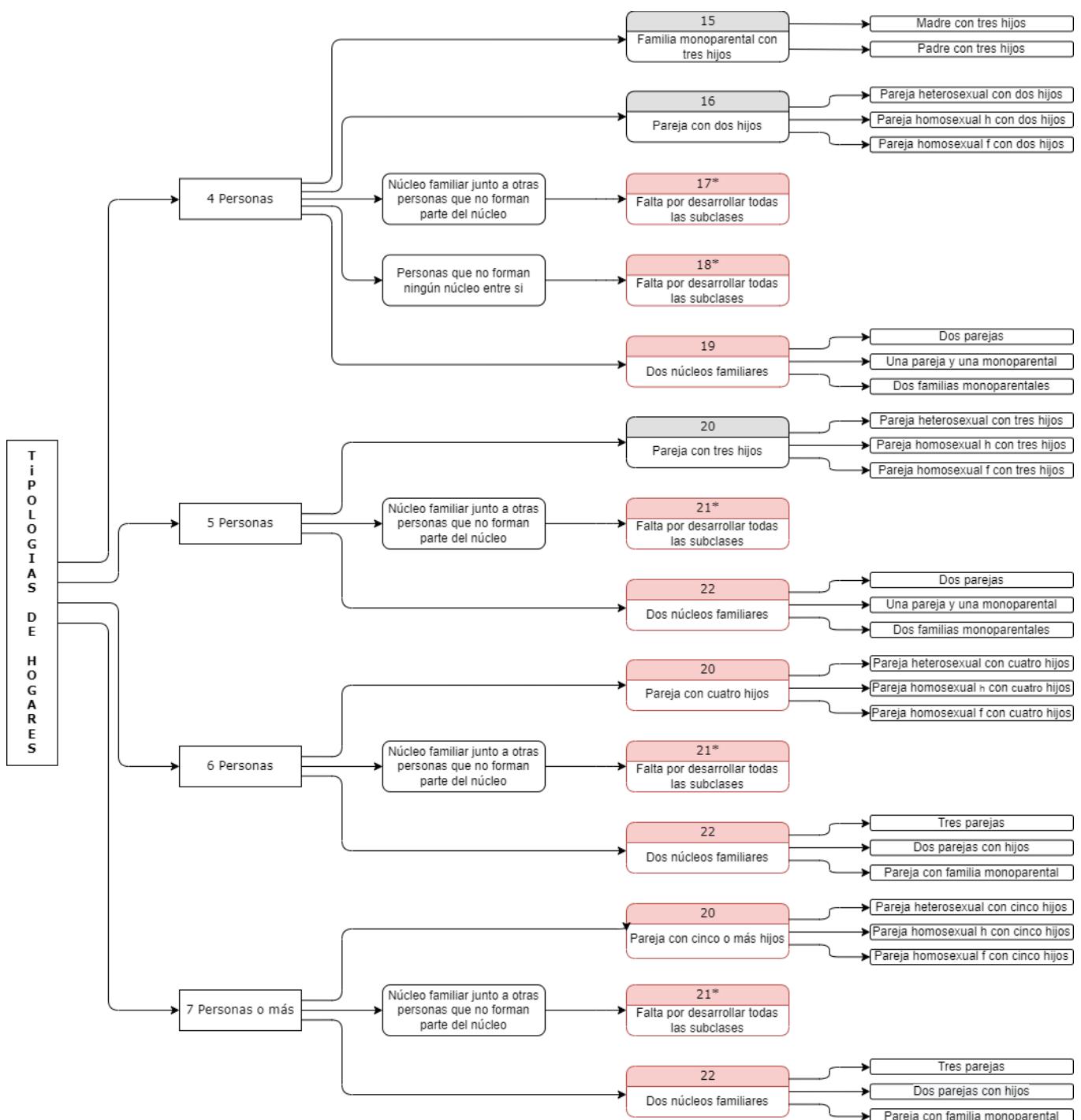


Figura 19. Diagrama tipologías de hogar

La implementación de las clases identificadas en el estudio no se llegó a completar, pues algunas de las clases que si fueron identificadas y verificadas, no se incluyeron en el código. En el caso de las tipologías de las familias más numerosas solo se han llegado a implementar las tipologías base, mostrándose en rojo en el diagrama todos los tipos de hogares que faltan por introducir.

Para poderse centrar en las siguientes tareas del sintetizador, se deja la identificación e implementación de las clases no incluidas para una segunda versión del programa.

3.2.5.3. Estructura y funcionamiento

Tras el estudio inicial de los factores a tener en cuenta e identificadas las tipologías a implementar, se empezaron a recopilar datos estadísticos mediante las tablas del INE o mediante consultas a los microdatos, con los que poder generar las familias. Una vez obtenida una base sólida de datos se empezó con la programación, aunque a medida que se avanzaba el desarrollo del programa, se identificaban nuevos problemas y características de las familias que requerían de nuevas estadísticas. Como la metodología ágil elegida para el proyecto es incremental, se iban estableciendo tipologías a implementar como objetivos, por lo que la búsqueda de nuevos datos se dio de forma pareja a la programación.

El Familiador está basado en estadística y se centra en tomar elecciones sobre los datos estadísticos de Leganés. Para poder escoger entre las distintas características o entre las distintas tipologías, el programa hace uso de un método que se explicará más adelante, pudiendo escoger así variables como la edad, el sexo o la situación laboral. Este primer módulo ha sido creando a partir del censo real de Leganés y de todos los datos estadísticos que se han conseguido reunir, consiguiendo como resultado una población sintética apta para simular.

La estructura de la primera parte del programa se puede observar en el diagrama de la Figura 20. El programa se compone de dos bucles principales, el primero sirve para cambiar de distrito y el segundo, que se encuentra dentro del primero, es el que genera las familias. El Familiador empieza con la inicialización de las variables para, a continuación, adentrarse en el primer bucle con el primer distrito. Se abre el fichero csv del distrito, se lee el censo y se guarda la población en dos listas, una de hombres y otra de mujeres, donde la posición de la lista indica la edad. A la hora de crear las familias, el Familiador distinguirá entre niños y adultos por medio de dos franjas de edad, de 0 a 24 se considerarán niños y de 25 a 95 se considerarán adultos, todas las personas con más de 94 años quedan agrupadas en el 95. Así, se consigue clasificar a las personas por sexo y edad.

Posteriormente, se lee el catastro del primer distrito de otro fichero csv, obteniendo los metros cuadrados y las coordenadas x e y de cada casa. Los metros cuadrados sirven para dividir las casas de Leganés en tres tamaños: grande, mediano y pequeño. En el caso de Leganés, al ser una “ciudad dormitorio”, el parque inmobiliario es más económico que en Madrid y no contiene la cantidad de pisos pequeños como estudios o lofts, que abundan en la capital. Por esta razón, el rango de tamaño de las casas pequeñas se ha aumentado para abarcar un número considerable de casa en esta lista, se establece por tanto, un tamaño máximo de 60 m² para las casas pequeñas, el tamaño de las casas medianas se establece entre 60 y 100 m² y por último las casas grandes son aquellas que cuentan con más de 100 m².

Una vez leídos y guardados censo y catastro, se inicia con la generación de las familias, que funciona con las estadísticas recogidas en porcentajes y que como se puede observar en el diagrama de la Figura 20, se obtienen del INE. La generación de familias se realiza mediante el segundo bucle, que va a ir creando familias constantemente hasta que no queden más adultos con los que generar nuevos hogares, momento en el cual saldrá de este segundo bucle, para iniciar con el siguiente distrito, volviendo a leer y guardar el censo y el catastro del siguiente distrito.

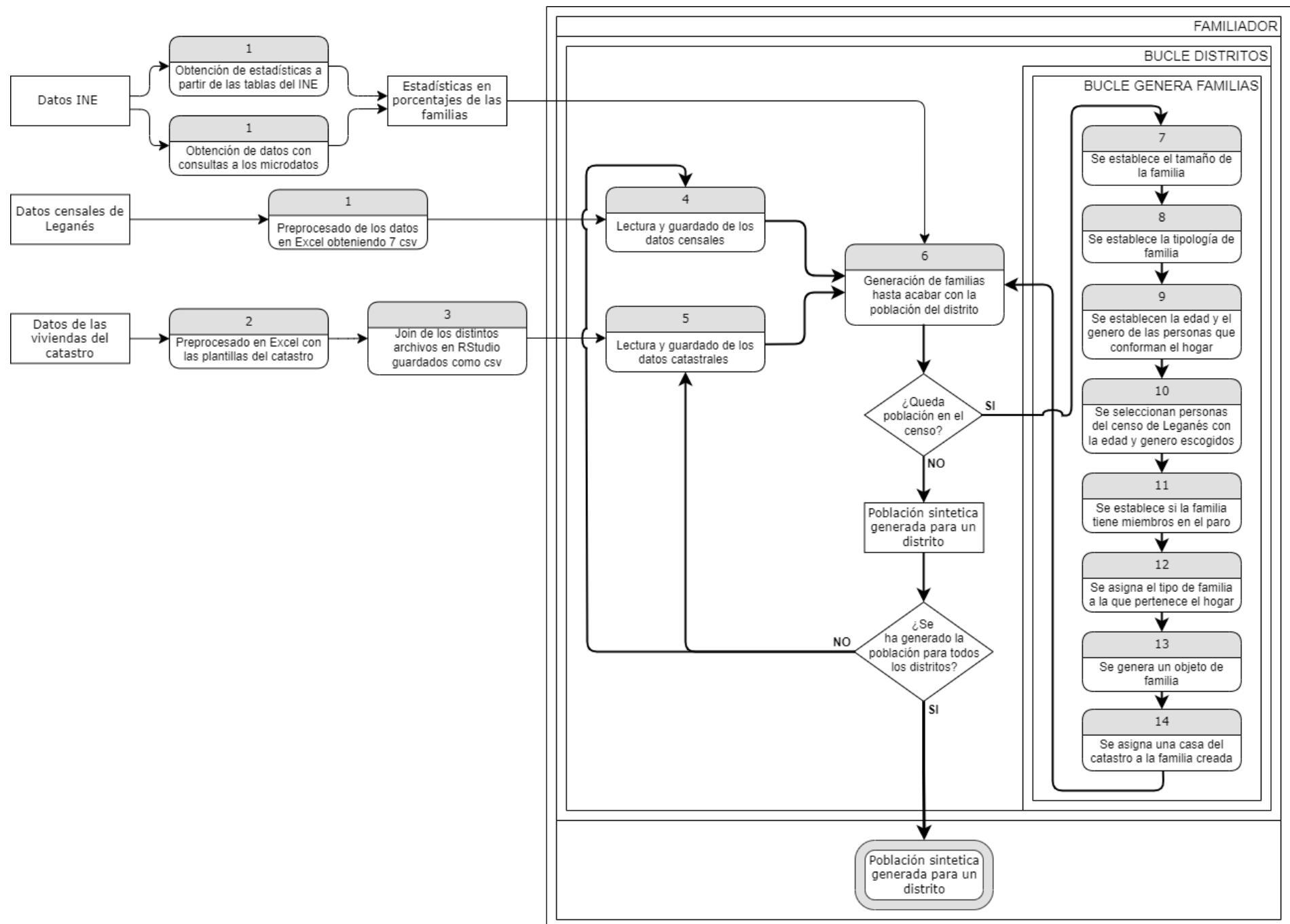


Figura 20. Diagrama del Familiador

Para tomar decisiones dentro del bucle de creación de familias, se utiliza siempre el mismo método. Primero tienen que estar definidas las categorías entre las que hay que decidir, como, por ejemplo, para la decisión del número de personas que componen el hogar, se tienen las categorías de casas compuestas por 1, 2, 3, 4, 5, 6 y 7 o más personas. A continuación, a partir de los datos del INE se obtienen los porcentajes de aparición de cada una de las casuísticas. Para adaptar estos porcentajes al programa, se crean unos rangos de aparición de las categorías. El intervalo de la primera categoría comienza en 0 y llega hasta el porcentaje de aparición de esta primera categoría, mientras que el rango de la segunda categoría comienza por la cifra porcentual por la que termina el primero rango y terminará en la cifra que se obtiene como resultado de sumarle a este último porcentaje, el porcentaje de aparición de esta segunda categoría. Este proceso se repite para el resto de las categorías hasta llegar a la última, en la que el rango llega hasta el 100%. Continuando con el ejemplo de la selección de la cantidad de personas que componen el hogar, se tiene la Tabla 4 donde se pueden observar cómo quedarían los rangos en este caso.

Una vez definidas las categorías y calculados los rangos, estos intervalos se introducen en el código en distintos condicionales. Pudiendo escoger entre unos y otros gracias a un generador de números aleatorios que produce cifras comprendidas entre 0 y 100 con un decimal, que se generan previamente a los condicionales.

categorías	1 persona	2 personas	3 personas	4 personas	5 personas	6 personas	7 o más personas
Porcentajes originales	25.7%	30.2%	20.1%	17.6%	3.9%	1.5%	1%
Rangos de porcentajes	0% a 25.6%	25.7% a 55.8%	55.9% a 75.9%	76% a 93.5%	93.6% a 97.4%	97.5% a 98.9%	99% a 100%

Tabla 4. Agrupación de estadísticas por rangos

El bucle que genera las familias se compone de una serie de pasos en los que se van a ir tomando distintas decisiones en cada uno de los campos que definen una familia. Decisiones que se realizan siempre mediante el método del número aleatorio explicado anteriormente, que van a ir configurando la familia en base a las estadísticas reales de la población. Los pasos por los que se compone este bucle son los siguientes:

1. El primer paso dentro del bucle genera familias consiste en decidir el tamaño de la familia a generar.
2. En el segundo paso se establece la tipología de la familia, escogiendo entre las tipologías existentes para el tamaño de la familia previamente elegido. Una vez escogida la tipología, pero antes de crear la familia, se comprueba que el censo contenga las personas exigidas por la tipología seleccionada, por ejemplo, para crear una familia monoparental se comprueba que queden niños disponibles. En caso de que no queden disponibles los tipos de personas especificadas, no se creará la familia, saliendo del bucle para volver a iniciar el proceso desde el principio.
3. El tercer paso consiste en establecer la edad y el género de las personas de la familia, este paso difiere en gran medida entre las distintas tipologías, dado que cada tipología tiene sus características específicas y contemplara distintas problemáticas que en otras tipologías pueden no ser necesarias. Un claro ejemplo de esto se puede observar en las

limitaciones por edad según la relación de parentesco entre las personas que componen la familia, como que una mujer de 80 años no pueda tener un hijo recién nacido. Todas estas diferencias entre tipologías a la hora de designar la edad y el género de las personas se comentarán más adelante en el siguiente apartado.

4. Una vez seleccionadas la edad y el género de las personas en el anterior paso, en este cuarto paso se van a seleccionar una a una las personas designadas de las listas del censo. En caso de que el censo no contenga personas de la edad deseada, se buscarán personas ligeramente mayores de forma recursiva, pero siempre diferenciando entre niños y adultos.
5. En el quinto paso se va a definir la situación laboral de la familia que se está generando, utilizando para ello unas estadísticas de paro específicas de cada tipología, obtenidas mediante consultas a los microdatos del INE, que fueron realizadas por Álvaro Delgado [56].
6. En el sexto paso se va a asignar al hogar una de las ocho tipologías que identifica el tipo de familia según la movilidad, las cuales fueron creadas para conectar las tipologías específicas del Familiador con la familias reales de la encuesta de movilidad, que se utilizarán en el siguiente módulo del programa.
7. Tomadas ya todas las decisiones se pasa a generar el objeto de la familia en cuestión, pasando todas las variables al constructor, el objeto de familia contiene los siguientes atributos:
 - a. Id: Identificador de la familia seleccionada
 - b. Personas: Lista que incluye el id, edad y genero de todas las personas que componen la familia.
 - c. Casa: coordenadas de la casa que se le ha asignado a la familia, que se obtienen en el paso 8
 - d. Tipo de familia: Número que identifica el tipo de familia según la movilidad de ésta, que fue asignado en el paso 6 y que será necesario para asignar los planes en el siguiente módulo.
8. Por último, en este paso se asigna una casa a la familia, para ello se utilizan unas estadísticas que varían en función del número de personas de las que se compone el hogar, asignando una casa pequeña mediana o grande, que se añade al objeto de la familia utilizando las coordenadas x e y. Dado que el número de familias que se generan es mayor al número de casas de Leganés que se obtuvieron del catastro, en el caso de que se vacíen las listas de casas, estas se rellenan con las mismas casas, pues previamente se generó una copia de las listas de casas iniciales.

Una vez creada la población de todos los distritos del municipio, tal y como se ve en el diagrama de la Figura 20. Se saldrá del primer bucle obteniendo la población sintética completa de Leganés, que será pasada al siguiente modulo del programa, es decir, el Planificador.

3.2.5.4. Selección del género y edad: características familiares

Familias unipersonales

La creación de esta tipología se inicia generando dos números aleatorios para designar cual es la franja de edad de la persona, con un primer rango de entre 18-25, que irá aumentará de 5 en 5 para los siguientes rangos y el género de la persona escogida. Si se diera el caso de que solo queden personas adultas de uno de los géneros, el número aleatorio creado para ello no se tomará en cuenta.

El género que predomina en los hogares unipersonales varía durante las distintas etapas de la vida, hecho que se ha podido confirmar con los datos del INE. Se ha constatado que durante la juventud las mujeres suelen ser las que primero se independizan, sin embargo, durante la vida adulta pasa lo contrario, son los hombres los que en mayores porcentajes viven solos. Por último, en la vejez, al tener las mujeres una mayor esperanza de vida, estas personas mayores enviudan y suelen vivir solas, por lo que predominan las viviendas unipersonales de este género.

Familias monoparentales

Los datos recogidos del INE mostraban que las familias monoparentales adolescentes representan unos porcentajes ínfimos por lo que no se han llegado a incluir a madres o padres por debajo de 25 años. Para generar esta tipología se comienza obteniendo la edad del padre o madre mediante un número aleatorio, una vez elegida la edad del progenitor, se continua con la del primer hijo, esto puede ser un problema. Debido a que el programa considera que las personas de entre 0 a 24 años son niños, si la edad del progenitor es baja y la del hijo alta, se podrían obtener dos edades cercanas, y dado que las madres solo pueden tener hijos a partir de cierta edad, la familia resultado no sería realista. Para solucionar esto, se establece que las mujeres pueden tener hijos a partir de los 15 años y a la hora de asignar cuantos años tiene primer hijo se debe tener en cuenta que, si la edad de la madre es inferior a 39 años, la edad del hijo estará restringida para que este no pueda haber nacido antes de que su progenitor tuviera 15 años.

Parejas

Para hallar las parejas se estuvo estudiando la diferencia de edad entre los dos miembros que la forman, buscando información se encontraron algunos estudios como “*La similitud entre los componentes de las parejas jóvenes en España en la primera década del siglo XXI ¿Cada vez más iguales?*” [58], que sirvieron para entender cómo han ido evolucionando las parejas españolas, de un pasado en el que el modelo de pareja era muy rígido, hasta la situación actual donde la pareja tradicional se ha diluido para dar paso a otros tipos de pareja donde ya no solo se encuentra la pareja heterosexual y donde la diferencia de edad en la pareja ha cambiado. La homogamia etárea es el término que se utiliza en sociología para hacer referencia a la diferencia de edad en las parejas, cuando ambos miembros de la pareja tienen la misma edad se denomina homogamia, cuando existe una diferencia de edad en la pareja existen dos términos, al tener mayor edad se denomina hipergamia y al tener menor edad se denomina hipogamia.

Por lo tanto, para poder reflejar correctamente la diferencia de edad se tuvo que realizar un estudio a base de consultas de los microdatos del INE que permitieron extraer las estadísticas necesarias que utilizaría el Familiador, obteniendo la Tabla 5 para las parejas de la comunidad de Madrid:

Diferencia de edad en la pareja según edad de la mujer							
Edad	n	N	Homogamia	Hipogamia moderada	Hipergamia moderada	Hipogamia elevada	Hipergamia elevada
20 - 24	81	66.597	19,16	0,00	11,59	0,00	69,25
25 - 29	230	234.153	23,73	3,22	7,68	1,46	63,91
30 - 34	695	526.012	29,26	3,62	11,16	7,38	48,59
35 - 39	1.581	653.667	32,03	4,14	12,05	8,30	43,48
40 - 44	2.071	815.098	34,80	4,65	12,94	9,22	38,38
45 - 49	1.082	808.754	32,51	4,93	10,63	10,88	41,05
50 - 54	1.028	709.955	34,88	4,39	11,57	8,33	40,84
55 - 59	829	598.257	32,86	4,12	12,01	10,01	41,00
60 - 64	777	489.693	33,66	3,98	15,07	6,38	40,91
65 - 69	697	427.858	29,53	3,65	12,52	6,08	48,22
70 - 74	562	337.338	32,67	4,10	11,39	8,72	43,12
75 - 79	459	268.378	34,19	3,75	13,72	3,48	44,86
80 - 84	248	140.074	32,83	2,60	13,91	7,12	43,54
85 - 89	115	65.803	36,70	12,02	17,54	15,78	17,96
90 - 94	18	10.129	51,23	16,51	9,64	11,65	10,97

Tabla 5. Diferencia de edad en la pareja

Por lo tanto, con los datos obtenidos sobre la diferencia de edad junto con los datos de la edad de las personas que viven en pareja, se genera este tipo de hogar. Esta tipología difiere en el caso de ser una pareja con hijos o sin ellos.

- En el caso de ser una pareja sin hijos, se utilizan unas estadísticas extraídas del INE sobre el rango de edad de las mujeres que viven en pareja. Estadísticas con las que se calcula el rango de edad de la mujer haciendo uso del número aleatorio
- En el caso de ser una pareja con hijos, dado que la edad de la madre viene condicionada por la del hijo, para obtener resultados más realistas se han obtenido estadísticas de la edad en la que las mujeres se quedan embarazadas de su primer hijo. Por lo cual, primero se calcula que edad tenía la mujer cuando tuvo a su primer hijo, para luego con un número aleatorio obtener la edad que tiene el hijo actualmente, pudiendo situarse dentro del rango 0 y 24 años, para finalmente obtener la edad actual de la madre, que se obtiene de la suma de la edad del primer hijo más la edad a la que lo tuvo.

Una vez seleccionada del censo y concretada la edad de la mujer, por cualquiera de los dos métodos, se pasa a decidir la diferencia de edad en la pareja. Los porcentajes de esto varían en función del rango de edad de ella, teniendo que elegir entre los siguientes tipos:

- Homogamia: Ambos tienen la misma edad o él tiene un año más.
- Hipogamia moderada: Ella tiene un año más que él
- Hipogamia elevada: Ella es mucho mayor que él
- Hipergamia moderada: Él tiene 2 o 3 años más que ella.
- Hipergamia elevada: Él es mucho mayor que ella

Cuanto mayor sea la diferencia de edad en la pareja, menores son las probabilidades de que exista la pareja. Por ello, para los casos en los que él o ella sean mucho mayores que su pareja, se hace una distribución de probabilidades entre los 20 años siguientes, donde una diferencia de edad menor será la ocurrencia más probable y donde a medida que vaya aumentando la diferencia de edad la probabilidad se irá reduciendo, restando un pequeño porcentaje equivalente por cada año. Para los casos de personas jóvenes y personas ancianas se ha controlado que las parejas no sean niños ni personas de más de 95 años que el programa no contempla, puesto que están contenidas todas ella en la posición 95 del censo.

Por último, el programa decide si la pareja es homosexual o heterosexual, comprobando que para cada uno de los casos queden las personas necesarias en el censo. En caso de haberse generado una pareja homosexual se habría tomado a un miembro de la pareja por la mujer y a otro por el hombre, hecho que no afectaría al cómputo global.

Personas que no forman ningún núcleo familiar entre sí

Dentro de este tipo más general de familia se han llegado a desarrollar dos tipologías, la de un par de hermanos convivientes en la misma casa y otra compuesta por un abuelo con su nieto.

Para los hogares de hermanos, primero se calcula si ambos hermanos son del mismo género o no mediante el número aleatorio, según la estadísticas de Madrid las parejas mixtas son mayoría. A continuación, se calcula el rango de edad del hogar y al igual que como se hacía con la diferencia de edad en las parejas, en esta tipología se hace de una forma una forma parecida. Se obtiene la edad de uno de los miembros del hogar en vez de la madre, y a partir de esta edad se obtiene la diferencia de edad con su hermano, teniendo los siguientes intervalos:

- Diferencia de edad de 0 a 2 años
- Diferencia de edad de 3 a 5 años
- Diferencia de edad de 6 a 10 años
- Diferencia de edad de más de 10 años

Escogida una diferencia de edad, se calcula si esta es por encima o por debajo de la edad del hermano ya concretado, asegurándose que el hermano no sea un niño pequeño.

Para los hogares compuesto por un abuelo con su nieto, primero se decide el género del adulto y del nieto, para después obtener sus edades mediante unas estadísticas obtenidas a través de consultas a los microdatos del INE. Primero se calcula cuantos años tiene el nieto, para luego en base a su edad obtener la edad del abuelo y así evitar que ambos tengan una edad pareja.

Se estableció que entre ambos debe haber como mínimo 40 años de diferencia de edad, lo que obliga a que como mínimo el abuelo tuviera un hijo con 20 años, para que luego su hijo tuviera

a su nieto también con 20. Por lo tanto, una vez elegida la edad del nieto se bloquean las posibles edades para el abuelo que no cumplan esta diferencia de 40 años.

Segundos hijos

Tanto para parejas como para familias monoparentales que tengan 2 hijos o más, se tiene que calcular la brecha entre un nacimiento y otro, esta diferencia de edad entre un hijo y el siguiente sigue unas probabilidades estadísticas. Por ejemplo, las familias que tienen muchos hijos suelen tenerlos muy seguidamente, por otra parte, cuanto mayor tiempo pase desde el nacimiento del último hijo, menor será la probabilidad de traer un nuevo hermano a la familia. Por ello, para todas las tipologías de familias con más de un hijo se llama a una función “siguenteshijos”, la cual calcula la edad de los hijos restantes a partir de la edad del primero, aplicando unas estadísticas del INE de las que se obtuvieron los siguiente categorías entre las que elegir:

- Tienen la misma edad, esto ocurre si son mellizos o gemelos
- Diferencia de edad de 1 año
- Diferencia de edad de 2 años
- Diferencia de edad de 3 a 9 años
- Diferencia de edad de más de 9 años

Al igual que con la diferencia de edad en la pareja, para los dos últimos casos se hace la misma distribución de probabilidades, donde la menor diferencia de edad es el caso más probable.

El cálculo de la homogamia etárea se hace hijo a hijo a partir de la edad del primero, es decir con la edad del primer hijo se calcula la diferencia de edad con el segundo, para luego a partir de la edad del segundo calcular la edad del tercero, razón por la cual los hijos posteriores deben tener como mínimo la misma edad y nunca inferior. Cabe comentar que este método se podría mejorar si se hiciera un estudio más completo mediante consultas, pues los rangos de diferencia de edad son poco exhaustivos y están calculados de hijo a hijo, cuando en realidad estos porcentajes dependen del número de hijos que tiene una familia y se deberían observar en conjunto, elaborando distintas probabilidades para cada tamaño de familia.

3.2.5.5. Mejoras

La generación de una población es una tarea compleja, pues el número de factores que influyen en la composición de las familias es elevado. Por otra parte, aunque las tipologías de hogar más comunes son mayoritarias, la composición del resto de hogares es muy heterogénea y compleja y, por tanto, difícilmente representable.

Al centrarse este trabajo en la creación de una guía completa para la preparación de la simulación basada en agentes, no se han podido abarcar todos los factores, tipologías y mejoras de la interfaz de usuario deseados, continuando con el siguiente modulo y obviando muchos puntos de mejora en este primer prototipo, pero siendo consciente de sus limitaciones actuales. Sin embargo, todas estas ideas de mejora surgidas alrededor del desarrollo del Familiador van a quedar reflejadas en esta sección, para facilitar el desarrollo de un segundo prototipo del programa a otros desarrolladores interesados en el proyecto.

Especialización por división administrativa

La versión actual del Familiador tiene unas estadísticas fijas para el conjunto de Leganés y no se hace una distinción de la composición de las familias por distritos. Aquí juega un papel importante la falta de datos a bajo nivel. La diferencia entre distritos en relación con la edad media de los ciudadanos, el nivel económico y el nivel educacional es considerable y debería de

tenerse en cuenta. Para hacerlo, se podrían utilizar como referencia los datos del censo de 2021 cuando éste se publique, pues los microdatos de éste contendrán parámetros a nivel de distrito con los que se podría abordar esta diferenciación. Para precisar mejor la población de Leganés, en lugar de realizar esta diferenciación por distritos se podría analizar por barrios, generando distintas estadísticas para cada uno de ellos.

Factores

Integrar en el software factores clave que no se llegaron a introducir en esta primera versión, podría constituir una mejora importante de la población resultado. Como, por ejemplo:

- La renta per cápita: Como ya se mencionó en el apartado de factores, este factor correlaciona significativamente con la composición de los hogares. Para introducir este factor, se podría implementar una nueva dimensión en la que se tenga en cuenta la renta de las personas, previa a la asignación de una tipología.
- El nivel de estudios de las personas: La posesión de mayores estudios retrasa el acceso al mundo laboral y la formación de una familia. Esta nueva dimensión modificaría la creación de familias jóvenes en función de su nivel de estudios.
- La nacionalidad: Uno de los subfactores que incluye hace referencia a la cultura de origen y la adaptación e integración a la cultura local. Éste sería un factor complejo de introducir, debido a que no suele haber registros referidos a la adaptación cultural.

Para introducir estas mejoras primero habría que establecer diferencias dentro de las tipologías, analizando en qué sentido y en qué grado están afectan los estudios o los ingresos a la aparición de estas familia, para luego al trabajar en los distintos distritos o barrios, aplicarles estas diferenciaciones según la renta o los estudios medios a cada uno de ellos.

Tipologías

Uno de los objetivos primordiales antes de agregar una mayor complejidad al algoritmo con la introducción de nuevos factores, consistiría en continuar con el estudio sociológico con la identificación e implementación de las tipologías de hogar, para que, aunque se generé la población de Leganés a nivel de todo el municipio, sin tener en cuenta las diferencias entre distritos o barrios, esta población contenga todos los tipos de hogares existentes.

Mejora de las tipologías implementadas

Muchas de las tipologías implementadas admiten distintas mejoras, algunas de éstas ya comentadas en el apartado anterior. Mejoras como la de realizar un nuevo estudio acerca de la diferencia de edad entre hermanos, o como mejorar el estudio de la edad de embarazo de las mujeres según el número de hijos que tuvieron finalmente.

Por otra parte, la situación laboral de los miembros del hogar que se introdujo en todas las tipologías se implementó como estadística general para cada tipo de hogar, sin hacer distinciones dentro de cada tipología. Esta variable se debería mejorar haciendo una diferenciación por edad o por estudios, creando distintos porcentajes de empleabilidad según las características de la familia.

Interfaz de usuario

Por último, para mejorar la usabilidad del Familiador se podría generar una pequeña interfaz de usuario, en la que poder introducir las estadísticas de la población requeridas, de modo que sea el mismo programa el que sustituya automáticamente los porcentajes actuales por los nuevos, ya que, en la actualidad, esta operación tan solo se puede hacer de forma manual. Al menos, se

debería fijar la estructura en la que se tendrían que introducir las estadísticas en un fichero csv para que al pasarle este fichero csv al Familiador, lo lea y modifique los porcentajes correspondientes. Esta modificación de porcentajes podría admitir que algunos de los campos estadísticos se queden sin llenar, para los casos en que las estadísticas obviadas no sean esenciales. Por ejemplo, en el caso de que no se tengan las estadísticas de una tipología, se podría obviar la ejecución de esa parte si el usuario no agrega tal estadística en la interfaz del programa.

3.2.6 Generación de los planes: El Planificador

Una vez generada la población sintética con el Familiador, el siguiente paso consiste en asignar a estas familias sintéticas una serie de planes iniciales que utilizaran los agentes de la simulación para desplazarse por el mapa de Leganés. Planes iniciales que tras las sucesivas simulaciones se van a ir optimizando, pero que necesitan una base inicial sobre la que partir.

Al igual que ocurría con la parte de la generación de población, todos los estudios consultados sobre la generación de planes diarios no estudian la movilidad a través de la sociología, si no que generan los planes a través de planes reales. Por lo tanto, para estudiar la movilidad a través de un análisis de las estructuras sociales, de la misma manera que se hizo con la creación de los hogares, habría que estudiar los factores que modelan y condicionan esta movilidad. Algunos de estos factores ya fueron comentados en el apartado del Familiador, puesto que afectaban tanto a la composición de los hogares como a su movilidad. No obstante, otros factores solo afectan a la forma en que se mueven las personas, por lo que para poder representar fielmente los planes diarios de movilidad habría que volver a estudiar los factores ya identificados e identificar y analizar nuevos factores, como, por ejemplo, habría que considerar como afecta una calle a la movilidad que se desarrolla por ella (si la calle estrecha o ancha, si es una calle larga y conecta con muchas otras, si es una calle comercial o no, etc.) Todo esto se debería analizar, haciendo uso de conteos de vehículo y personas, tal y como hizo el ayuntamiento de Leganés en un estudio de movilidad sostenible [59]. Otro posible factor a analizar podría ser el estudio acerca de cómo las obras y los atascos afectan a la movilidad, o estudiar como la fecha en que se dé la movilidad influye en los viajes (si es en días de diario o en el finde semana o incluso en días festivos como navidad). La climatología sería otro factor a estudiar (la lluvia hace que más personas cojan el coche para desplazarse) y un largo etcétera de factores.

Como se ha reflejado, la movilidad no depende únicamente de las ya complejas interrelaciones que se dan entre los miembros de la unidad familiar o entre los miembros del hogar con otras personas de fuera de este núcleo familiar. Definir la movilidad teniendo en cuenta las estructuras sociales es mucho más complejo que definir la composición de las familias también mediante estas estructuras. Esto es debido a que la movilidad es influída por una mayor cantidad de factores y a que las fuentes gubernamentales recogen un menor número de datos de movilidad por cuestiones de protección de datos. Bien es cierto que esta realidad está cambiando, dado que se están realizando estudios de movilidad de las personas a partir de datos del posicionamiento de los teléfonos móviles [60]. Este análisis lleva a la conclusión de que la movilidad de los ciudadanos es difícilmente reproducible basándose en datos sociales. Crear planes sintéticos realistas es mucho más complejo que hacerlo a partir de planes reales sacados de encuestas de viajes.

Finalmente, al ver que la complejidad y el trabajo necesario para realizar un primer estudio social había sido constatada en la generación de los hogares y vista la cantidad de factores distintos a estudiar, se decidió no elaborar los planes diarios a través de un estudio sociológico. Tomando por tanto un enfoque más tradicional para la generación de planes. Por razones prácticas se decidió priorizar en este punto el objetivo de crear una guía clara y funcional para los usuarios

que deseen realizar simulaciones con sistemas basados en agentes, frente a la búsqueda de un nuevo enfoque para la generación de poblaciones sintéticas. De haber continuado con el enfoque sociológico, no se podría haber terminado la generación de la población con sus planes diarios.

Creación de la planificación diaria a partir de planes reales

Para este segundo modulo del programa se va a partir de planes reales realizados por ciudadanos de una serie de poblaciones del sur de Madrid, obtenidos de la encuesta de movilidad del consorcio de transportes de Madrid (CRTM), que van a servir de base para la creación de nuevos planes, tal y como se hace en otros trabajos [47].

La generación de los planes diarios para las familias sintéticas utilizará la base de los planes reales obtenidos del CRTM, planes diarios a los que se eliminarán las coordenadas de las ubicaciones de origen y destino, pero que conservan el resto de atributos, como el tipo de transporte elegido para el trayecto, el tipo de actividad a realizar en destino o las horas a las que se realizan los desplazamientos. Por lo tanto, para poder asignar planes a las familias sintéticas creadas, se van a buscar familias reales que realicen una movilidad diaria similar. La similitud se establecerá en función de sus características, por ejemplo, de la edad, la planificación diaria de una pareja joven no tiene ninguna relación con la de una pareja de ancianos.

La asignación deberá realizarse miembro a miembro, buscando para cada persona del hogar sintético, una persona similar del hogar real, es decir, intentar asignar a la madre sintética, los planes de la madre real, continuando del mismo modo con su marido y con los niños. Cuanto mayor similitud exista entre las dos familias, mejores resultados se van a obtener. Ambas familias deberían tener una misma composición, un mismo nivel de estudios, o una misma situación laboral. Sin embargo, al trabajar con la elevada población de una ciudad, no es factible realizar este proceso de intercambio de planes buscando para cada familia sintética la familia real exacta. Por ello, se van a tener que crear unas tipologías más generales que agrupen los hogares por una serie de factores, teniendo en cuenta dos problemáticas. La primera referida a los datos disponibles, pues podría darse el caso de que la encuesta de movilidad no tuviera los parámetros necesarios para poder clasificar los hogares reales de la misma manera en que están clasificados los hogares sintéticos. Una segunda problemática viene dada por la cantidad de datos disponibles, ya que, si el número de ejemplos de casas reales no es muy alto y buscando una alta semejanza entre hogares, se crean unas clases muy restrictivas, se podrían llegar a tener clases sin apenas ejemplos. Por todo esto, se deberá buscar un punto intermedio entre similitud y practicidad, donde los datos de los que se disponga hagan posible la creación de un mayor o menor número de clases.

3.2.6.1. Clasificación de los hogares reales

La idea inicial para establecer la clasificación de los hogares reales fue utilizar las mismas tipologías de hogares usadas en el Familiador para crear la población sintética, sin embargo, esto presentaba varios problemas. El problema principal derivaba de que la encuesta del CRTM no contenía información sobre la relación de parentesco entre los convivientes de los hogares encuestados, esto dificultaba identificar los tipos de hogar, que con la escasa información disponible solo se podían intuir estas tipologías por medio de las edades y el género de las personas de las que se componía cada hogar. Además, como ya se comentó en el apartado de datos, aún sumando el número de personas reales encuestadas de los cuatro municipios escogidos, no se disponían de datos suficientes como para realizar una clasificación tan exhaustiva como la elaborada en el primer módulo, pues quedarían muchas clases con escasos ejemplos. Un último problema, a colación del anterior, viene de que para poder representar correctamente la movilidad habría que incluir previamente otros factores que no afectan a la

composición de las familias, pero si a la movilidad. Este hecho, obligaba a que las múltiples tipologías ya desarrolladas en el Familiador tuvieran a su vez que ser divididas por su situación laboral o por la presencia de niños en ellas, ya que estos dos factores afectan en gran medida a la movilidad.

La clasificación final desarrollada para las familias reales simplifica las tipologías identificadas en el Familiador, pero agrega los dos factores anteriormente mencionados. Los factores diferenciadores que se han tomado como referencia para esta nueva subdivisión son dicotómicos, puesto que solo tienen dos valores posibles y son los siguientes:

- Tamaño del hogar (unipersonal/multipersonal): Esta variable divide a los hogares unipersonales de los multipersonales, dado que la movilidad difiere ampliamente ambos tipos. En los hogares unipersonales las tareas de la casa se concentran todas en una persona, y ésta no depende de otros miembros para realizar sus planes. Por ejemplo, en los hogares multipersonales es frecuente que unos miembros tengan que esperar a otros para poder realizar alguna tarea, también es frecuente que un miembro recoja a otro, para así cuadrar sus horarios, o incluso que se dedique tiempo de atención hacia otro individuo de la familia.
- Situación laboral (empleado/desempleado): Esta variable es crucial en la división entre tipologías puesto que las personas empleadas tienen unas rutinas diarias predefinidas, como el desplazamiento al lugar de trabajo, su estancia en él y su posterior regreso al hogar. Lo que hace que gran parte del tiempo este cubierto, dejando menor espacio para realizar otros planes. Mientras que las personas desempleadas se caracterizan por un tipo de desplazamiento más errático, donde no hay una rutina tan establecida y donde los tipos de desplazamientos pueden variar significativamente entre un día y otro. Aunque se poseían datos suficientes, esta variable admitiría una mayor profundización, dado que el tipo de trabajo condiciona en gran medida los desplazamientos, mientras los trabajos de oficina no implican movimiento alguno, otros trabajos, como puede ser el de repartidor, si exigen una gran cantidad de desplazamientos. También se podrían destacar algunas movilidades especiales como las que realizan las amas de casa o los jubilados.
- Presencia de niños (si/no): También esta variable resulta de gran relevancia, pues los hijos son muy dependientes de las personas adultas e implica una gran diferenciación entre unos hogares y otros, el tener que llevarlos al colegio o el ir a pasear al parque o llevarlos a realizar actividades extraescolares van a modificar significativamente los planes de un hogar.
- Familia monoparental (si/no): Esta variable resulta una pequeña subdivisión de la anterior y viene dada porque establecer los planes en las familias monoparentales es más complicado, puesto que mientras la exigencia que requiere la atención de un niño no cambia, la repartición de tareas y planes entre los adultos no se da, razón por la cual este tipo de familias están más sobrecargadas. Por ellos se decidió separar las familias con niños en dos tipos.

Combinando los factores mencionados, se concretó la subdivisión de los hogares en 8 clases, que agrupaban los hogares que compartían una movilidad similar:

1. Unipersonal – Parado – sin niños
2. Unipersonal – Trabajando – sin niños

3. Multipersonal – Parado – sin niños
4. Multipersonal – Trabajando – sin niños
5. Multipersonal – Parado – con niños – monoparental
6. Multipersonal – Parado – con niños – no monoparental
7. Multipersonal – Trabajando – con niños – monoparental
8. Multipersonal – Trabajando – con niños – no monoparental

Asumir esta clasificación de los hogares reales obtenidos de la encuesta de movilidad, implicaba utilizar la misma clasificación en los hogares sintéticos generados en el Familiador, para así poder traspasar los planes diarios de familias reales a familias sintéticas semejantes que compartieran una movilidad similar. Para hacer esto posible, a cada objeto de hogar del Familiador, se le agregó un atributo llamado “tipofamilia”, que indicaba a cuál de estas 8 clases pertenecía cada hogar. Para llenar esta variable, antes de generar el objeto, durante la creación de cada tipología se hallaba la clase a la que pertenecía el hogar, ya que se tenían los datos del tamaño del hogar, de la presencia de niños y de si se trataba de una familia monoparental o no, a esto se le agregaba la situación laboral de la familia, que se calculaba mediante estadísticas de empleo que fueron obtenidas por Álvaro Delgado a través de consultas a los microdatos del INE para cada una de las tipologías de hogar definidas.

3.2.6.2. Estructura y funcionamiento

Una vez establecida la clasificación de las familias reales y sintéticas en función de la movilidad, el siguiente paso consiste en realizar el traspaso de los planes de las familias reales a las familias sintéticas y agregar las nuevas ubicaciones a cada uno de los planes. Esta es la tarea que realiza el segundo módulo de *The Synthetic Population* que fue llamado Planificador.

En la Figura 21 se puede ver la estructura general del Planificador. Este módulo está subdividido en 4 submódulos que se reparten las distintas funciones que se tienen que realizar para poder asignar la actividad diaria a las personas generadas.

- 1. Creación del Ráster:** Este submódulo se encarga de generar el mapa de Leganés dividiéndolo en cuadrantes, al que agrega datos de los edificios del catastro para posteriormente poder dar nuevas coordenadas a los destinos de los planes reales.
- 2. Preprocesado de la población real:** Este submódulo obtiene de la encuesta del CRTM los datos de los planes diarios de familias reales, ordenándolos y estructurándolos para que puedan ser leídos por el siguiente módulo.
- 3. Traspaso de planes:** Este es el submódulo que realiza el traspaso de los planes de las familias reales a las familias sintéticas. Para ello, obtiene una familia sintética, a la que, tras comprobar su validez, le asigna una familia real que tenga una composición similar, escogiendo para cada persona sintética una persona real similar en edad y sexo, a continuación, se van pasando los planes persona a persona a excepción de las ubicaciones que se obtienen en el siguiente módulo. Tras recibir las nuevas ubicaciones, se escriben en el fichero XML de salida todas las personas de las que se compone el hogar junto con sus planificaciones diarias.

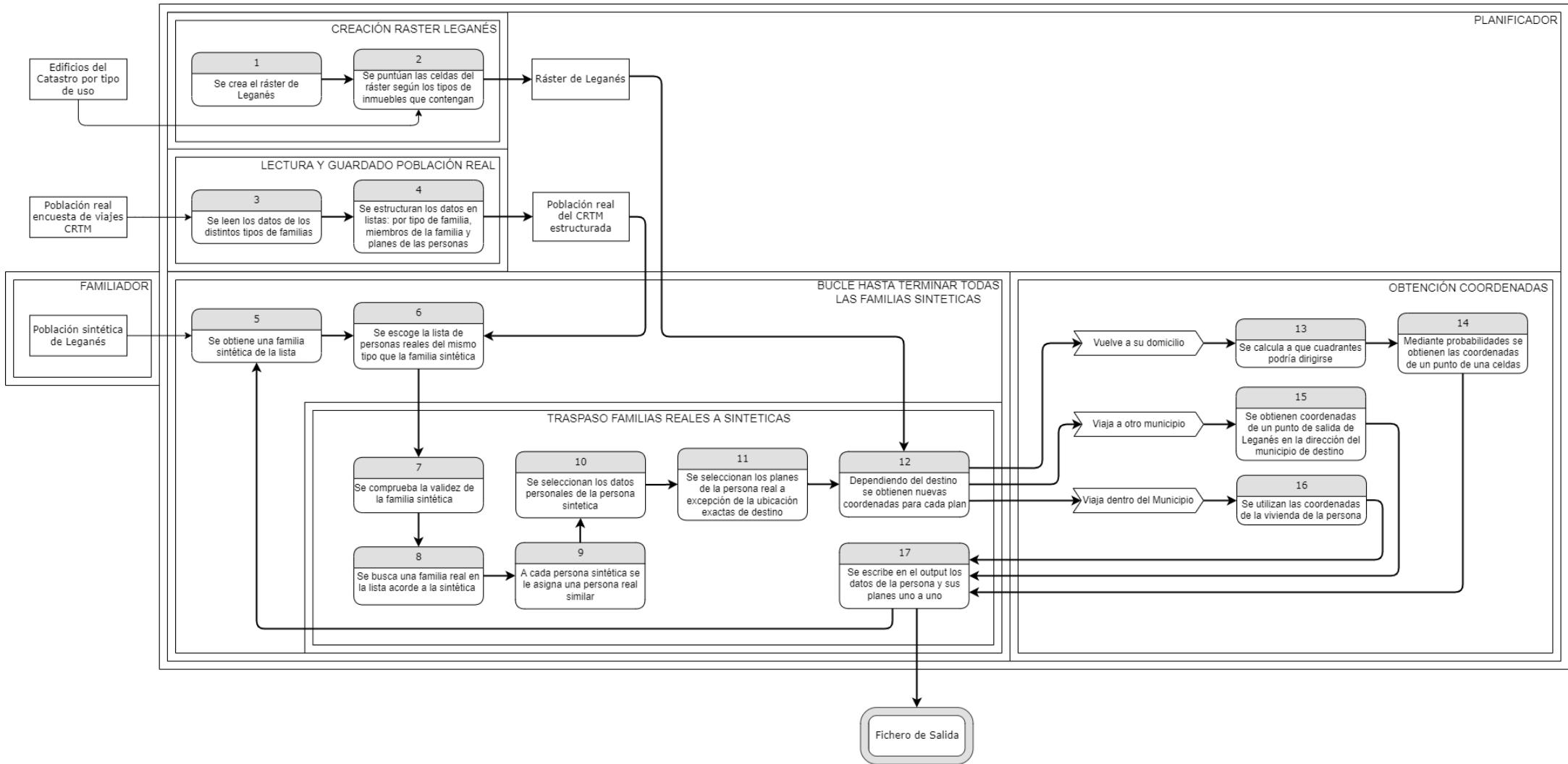


Figura 21. Diagrama del Planificador

4. Obtención de nuevas coordenadas: Este submódulo se encarga de completar los planes de las personas asignando nuevas coordenadas de destino, el cálculo de las nuevas coordenadas difiere según el lugar de destino del plan.

- a. Domicilio de la persona: Se devuelven las coordenadas de su casa.
- b. Fuera del municipio: Se devuelven las coordenadas de la carretera de salida de Leganés que se dirige al nuevo municipio.
- c. Dentro del municipio: Se calcula en función de la distancia y del tipo de actividad el cuadrante de Leganés al que se dirige, devolviendo las coordenadas de este.

El Planificador recibe una lista de objetos de familias sintéticas del Familiador e inicia el primer submódulo obteniendo el mapa de Leganés, para a continuación, obtener los planes desorganizados de las familias reales, ya que estos planes vienen uno tras otro. A través del segundo submódulo se estructura los planes, asignándolos a una persona, asignando las personas a una familia y finalmente asignando las familias a la lista de tipología de hogar a la que pertenecen.

Una vez se dispone del ráster, de los hogares reales y de los hogares sintéticos, el programa entra en un bucle que hará uso del tercer submódulo, para transferir los planes de unas familias a otras, este bucle se finaliza cuando se hayan procesado todas las familias sintéticas. Obteniendo como resultado la población sintética final con los planes diarios de movilidad incluidos y, listos para ser utilizado en Matsim.

3.2.6.3. Ráster de Leganés

Para poder asignar nuevos destinos a los planes que se desarrollan dentro del municipio es necesario crear un mapa de Leganés sobre el que realizar los cálculos, estos cálculos se realizarán en base a las distancias y a los tipos de actividad de los planes originales. Tomando como punto de partida las coordenadas de origen y aplicando sobre este punto la distancia recorrida en línea recta, se haya un nuevo punto que indicaría uno de los posibles destinos. Al aplicar esta misma distancia en el resto de direcciones, se pueden observar todos los posibles puntos de destino a los que una persona podría viajar. Para elegir un destino concreto, es decir, uno de los puntos hallados, habrá que ver cuáles de los destinos posibles tiene una mayor cantidad de infraestructuras que puedan servir al propósito del viaje.

Para elegir entre los puntos seleccionados como posibles destinos, se hace uso de las infraestructuras aledañas a cada uno de estos, pues estos puntos por si solos indicarían en el mejor de los casos, un inmueble. Se tiene que definir por tanto un área para todos los puntos, en la que se contabilicen las infraestructuras de cada tipo, para al comparar la cantidad de



Figura 22. Ráster de Leganés

infraestructuras aledañas a cada punto, se pueda elegir uno de ellos. Para hacer esto se va a zonificar Leganés, creando un ráster donde cada celda va a incluir una serie de inmuebles. A partir de las coordenadas de cada punto se podrá calcular la celda en que se encuentran, adquiriendo el listado de infraestructuras de la celda. De esta forma se podrá establecer los destinos de los planes de las personas en base a los inmuebles.

El mapa de Leganés con el que se va a trabajar se muestra en la Figura 22. El área seleccionada tiene 36 km² e incluye todas las zonas urbanizadas e infraestructuras importantes del municipio, a excepción de los terrenos al sur de la M-50 que son en su mayoría de uso agrícola y parques que quedan excluidos. Las principales carretera y circunvalaciones de Leganés han sido incluidas para que se pueda representar la movilidad correctamente, descartando la M-50 que, aunque está enmarcada dentro del término municipal, esta obligaba a agrandar el mapa para incluir zonas sin relevancia en la circulación intramunicipal que es la que se pretende simular en este trabajo.

Aplicando este procedimiento, Leganés quedaría dividido en celdas de 400x400 metros, teniendo un total de 225 celdas. Este tamaño de celda fue elegido buscando una distribución equilibrada en el número de infraestructuras por cuadrante. Elegir una subdivisión del mapa en celdas más pequeñas habría implicado un número muy elevado de ellas y una menor cantidad de inmuebles por celda. Por otro lado, una subdivisión mayor agruparía demasiados inmuebles y las coordenadas destino no serían representativas del tipo de actividad a desarrollar.

Tipos de actividades

Tal y como se comentó en el apartado de datos, el Catastro clasifica los inmuebles en función del tipo de uso que se les da, mostrándose en la tabla de Figura 16 todos los usos que utiliza el Catastro. Por ello, los inmuebles contenidos en cada celda del ráster estarían divididos por los tipos de uso del Catastro, sin embargo, para poder trabajar con Matsim, estas categorías de inmuebles fueron simplificadas en 7 clases, atendiendo al tipo de actividad por la que acuden las personas a ellos.

Por otro lado, los planes de las personas reales obtenidos de la encuesta de movilidad tenían asociado un motivo de viaje, por el que se desplazan a estos inmuebles. El CRTM clasificaba los planes por 12 motivos de viaje: casa, trabajo, gestión de trabajo, estudio, compras, médico, acompañamiento a otra persona, ocio, deporte/paseo, asunto personal, otro domicilio, otros. Al igual que con el Catastro, estos motivos fueron simplificados en las mismas 7 clases.

Para trabajar conjuntamente con el Catastro y con la encuesta del CRTM era necesario unificar ambas clasificaciones para así tener los planes organizados por los mismos tipos de actividad. Por otra parte, al trabajar con Matsim se suelen utilizar un número limitado de tipos de viajes, aunque no existen unos tipos específicos y se pueden agregar cuantos tipos se deseen, cada tipo de actividad que se quiera implementar tendrá que ser incluido en el fichero *config*, donde se especificarán sus parámetros de puntuación, para que luego Matsim sepa puntuar los planes tras cada iteración. Por todo esto y para facilitar la simulación, se simplificaron los viajes en 7 clases.

La Figura 23 contiene en la primera columna todos los motivos de viaje utilizados por la encuesta de movilidad, en la tercera columna se muestran todos los usos de los inmuebles identificados por el Catastro y por último, en la columna central se presentan los tipos de actividad asignados por el Planificador a los planes de las personas sintéticas, que más tarde serán usados por Matsim.

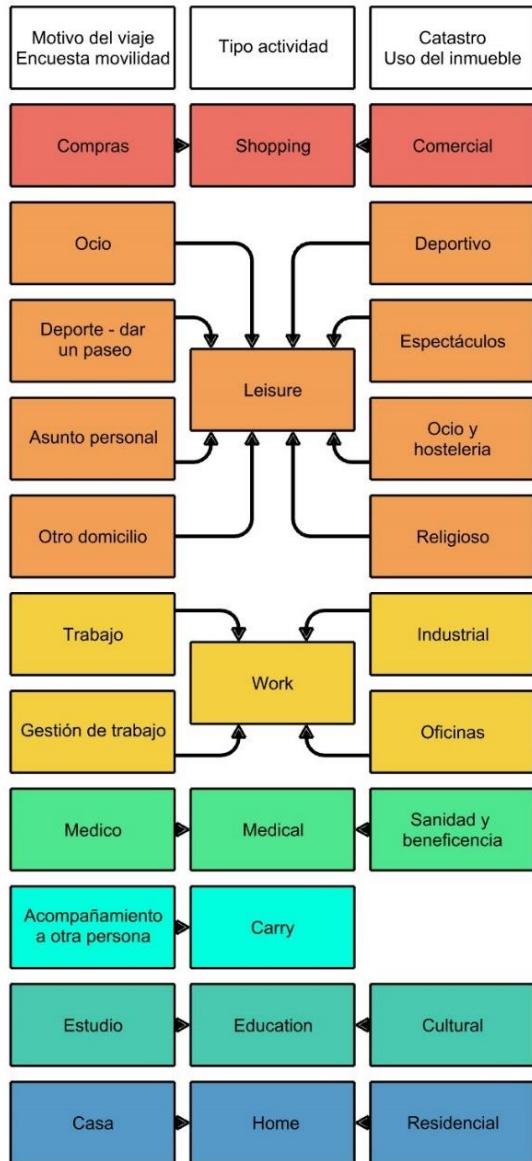


Figura 23. Conversión tipos de actividad

son necesarias para calcular los desplazamientos dentro del municipio, por lo que el tipo *home* no se contempla en el ráster, en el caso de *carry*, como se podía observar en el diagrama de los tipos de actividad, no tenía una equivalencia en el Catastro, ya que este tipo de actividad hace referencia a cuando una persona acompaña a otra para que esta realice una actividad, como acompañar a alguien al médico o como llevar a los hijos al colegio, por lo que no existen unos inmuebles donde se realice la actividad *carry*.

En esta misma Figura 23 se muestran mediante flechas las equivalencias entre los motivos de viaje de la encuesta de movilidad con los tipos de actividad asignados por el Planificador y las equivalencias de los usos de los inmuebles del Catastro, también con los tipos de actividad asignados por el Planificador.

Por lo tanto, cada celda del mapa de Leganés va a contener una serie de inmuebles que originalmente estaban clasificados según los usos del Catastro de la tercera columna, pero que serán reclasificados utilizando los tipos mostrados en la columna central. A modo de ejemplo, para ver los tipos de edificios y su clasificación se incluye a continuación en la Figura 24 la celda del ráster en la que se encuentra la Universidad Carlos III de Madrid.

Como se puede apreciar en la Figura 24, la celda elegida no contiene a la universidad en su totalidad, dado que hay partes de esta que se encuentran ubicadas en otras celdas del ráster. Aun así, se puede ver como al contener la mayoría de edificios de la universidad, el tipo de actividad *education* es la mayoritaria de la celda, la actividad *shopping* es el segundo tipo de actividad que más aparece, esto es debido a que la Avenida de la Universidad es una calle comercial. Por otro lado, los tipos de actividad *home* y *carry* no están puntuados. En el primer caso, esto se debe a que las casas ya fueron asignadas en el Familiador y no



Figura 24. Celda del ráster

Cabe comentar que la clasificación del Catastro tenía algunas limitaciones. Para empezar, cataloga a los colegios, universidad y demás edificios educativos como culturales, mezclándolos con museos, teatros y otros que podrían confundir los viajes. Además, no especifica claramente los tipos de inmuebles, solo agrupaba los inmuebles por sus usos, lo que impide poder definir la mayor afluencia a una celda concreta desde una perspectiva sociológica. Para solucionar estas limitaciones se ha optado por calcular la suma de los metros cuadrados por uso de los inmuebles, ya que era el único dato disponible para puntuar las celdas. Estas puntuaciones servirán de guía para elegir a una celda de entre todas las disponibles, como destino de un plan.

Mejora de la puntuación para un futuro

Un proyecto a futuro consistirá en mejorar la forma de puntuar los tipos de actividades, utilizando para ello la proyección social que se ha ido desarrollando durante todo este proyecto y es que, la no-linealidad típica de muchos fenómenos sociales, la aleatoriedad e irracionalidad humana hace que las decisiones de ir a un lugar o a otro sean muy complejas, debido a que son muchos los factores distintos que influyen en la elección de un destino. A modo de ejemplo se van a citar alguno de estos factores, la incidencia que tienen y algunas posibles modificaciones que mejoraría el cálculo de los planes de desplazamiento:

- **La edad:** Dentro de la categoría *leisure* hay muchos tipos de infraestructuras, un parque con columpios puede ser ideal para una familia con hijos, pero no así para una pareja joven, al igual que una discoteca va a ser frecuentada en su mayoría por jóvenes. En la categoría *education*, se podría distinguir a los colegios, institutos y universidades por edad, pues la afluencia a unos centros u otros depende del rango de edad de los estudiantes. Por ello, clasificar las infraestructuras por edad o por tipología del hogar haría que determinados inmuebles tuvieran mayores probabilidades de ser elegidos como destino para determinadas personas, creando categorías por edades se podrían mejorar los resultados del algoritmo.
- **El género:** Al igual que con la edad, el género puede ser determinante a la hora de elegir unos destinos u otros. Determinados negocios son mayormente visitados por un género, como por ejemplo los casinos, donde la mayoría de sus clientes suelen ser varones [61] o tiendas de ropa y centros de estética, donde se da el caso contrario. Pudiendo categorizar en ciertos porcentajes los posibles destinos por género.
- **El tiempo:** Dependiendo del día o de la hora, la elección de un destino u otro puede variar significativamente. Por poner un ejemplo, cuando una persona sale del trabajo en un día de diario, es mucho más probable que vaya a un bar con compañeros de trabajo, que ir a la bolera, un segundo ejemplo podría ser el de ir a un gimnasio después de la hora de comer, que es menos probable que a otras horas del día. Por ello, podría ser interesante realizar una clasificación de las infraestructuras en función de la fecha y hora en la que se produce la actividad, teniendo en cuenta las horas y días de afluencia máxima de los distintos lugares.
- **La economía:** Que un negocio no sea económicamente rentable, puede correlacionar con el hecho de que la afluencia de personas a éste haya descendido, esto podría ser discutible, pues hay negocios que consigan grandes márgenes de beneficios sin tener gran afluencia de personas, aun así, este factor debería ser candidato para ser estudiado en el futuro.
- **La meteorología:** Los destinos al aire libre, como terrazas o parques en días de lluvia o frío, tienden a tener una afluencia menor de personas. Esta misma variable produce en

los días de inclemencias un aumento en centros y locales cerrados. De ahí se deduce que una variable meteorológica debería ser incluida para analizar los destinos de los planes de desplazamiento.

Estos son solo algunos de los factores que se podrían implementar en versiones futuras del programa, profundizando en ellos se podría mejorar la elección de destinos actual. Sin embargo, para poder tener en cuenta muchos de estos factores, sería necesario conocer con exactitud los bienes inmuebles, consiguiendo datos que el Catastro no ofrece, pues únicamente muestra las tipologías de uso y los metros cuadrados.

Para realizar una asignación más exhaustiva, muchos proyectos han hecho uso de Open Street Map aprovechando que este tiene categorizados exhaustivamente los inmuebles como se puede ver en la [wiki](#) de OSM, sería posible obtener más datos de los inmuebles para realizar una mejor asignación de destinos a la población sintética, teniendo en cuenta así los factores anteriormente mencionados. Trabajos como los realizados en [62] y en [63].

Otra forma de abordar este problema en el caso de no disponer de los datos necesarios, consiste en generar de *motu proprio* los inmuebles, como hacen Zhuge et al. [44]. En este trabajo a partir de una pequeña muestra tomada de una encuesta de movilidad, se van generando las infraestructuras y se van posicionando en el mapa, creando así las infraestructuras de la ciudad.

Funcionamiento del primer submódulo

La creación del ráster del Leganés es básica para generar nuevos destinos y es el primer paso para la generación de los planes diarios, la estructura de este submódulo se puede ver en la Figura 25.

El proceso para la obtención del ráster comienza con el preprocesado previo al programa de los datos del Catastro, obteniendo las coordenadas y los metros cuadrados de los inmuebles de Leganés en un único archivo. El Planificador va a ir creando las celdas del ráster a partir de unas coordenadas iniciales con las que se calculan para cada celda las coordenadas que la delimitan, para cada celda se crea un objeto de celda, inicializando sus variables, estableciendo todos los tipos de actividades (*Shopping, Leisure, Work, Medical y Education*) a cero.

Una vez creadas todas las celdas que componen el mapa, se van leyendo uno a uno los inmuebles del catastro, calculando la celda en la que se sitúan a partir de sus coordenadas. Establecida ya la celda, se traduce el tipo de uso de inmueble del Catastro al tipo de actividad de Matsim a la que pertenece este inmueble, aumentando la puntuación de este tipo tanto como los metros cuadrados que tenga. Este proceso de clasificación de los inmuebles y de aumento de las puntuaciones de las celdas continúa hasta haber clasificado todos los edificios del Catastro, devolviendo el ráster de Leganés para los siguientes submódulos.

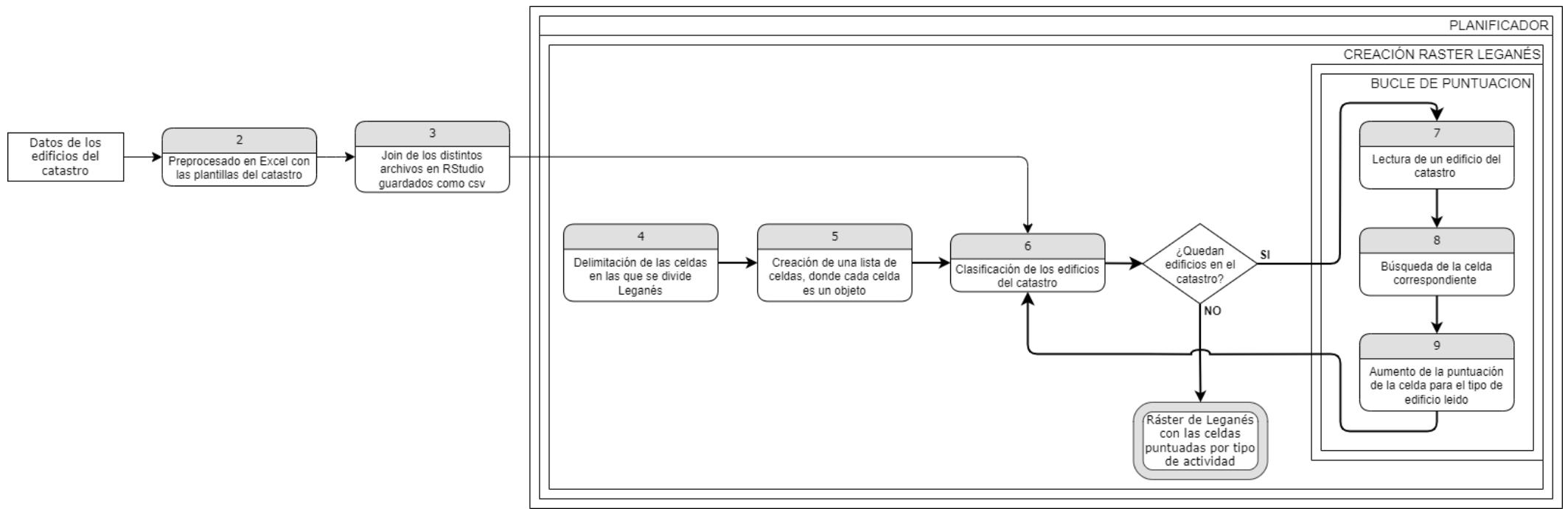


Figura 25. Diagrama submódulo 1

3.2.6.4. Preprocesado de la población real

Aunque ya se había realizado un amplio preprocesado de los planes obtenidos del CRTM previo al programa, que fue ampliamente desarrollado en el apartado 3.2.4.4. Datos de los desplazamientos de la encuesta de movilidad del CRTM. El segundo submódulo continua con el preprocesado de los datos dentro del propio Planificador. Esto es debido a que los planes recibidos no están agrupados, el programa recibe un fichero csv de una de las tipologías, donde están listados todos los planes, planes que contienen todos los datos, como por ejemplo la persona que los realiza o el hogar al que pertenece esta persona, etc. Por lo tanto, para asignar los planes reales esto se deben organizar primero, siguiendo la estructura que se muestra en la Figura 26.

Los planes recibidos son realizados por personas concretas. Para agruparlos se crea una lista para cada personas que contendrá todos los planes que realiza en el día. Al tener que asignar a un hogar sintético los planes de un hogar real, no bastará con agrupar los planes por persona, por lo que se crea una lista por cada familia que agrupará todas las listas de personas que conviven en el hogar, esto ya permitiría asignar todos los planes de una familia a otra. No obstante, como se han creado 8 tipologías de hogares para el traspaso de planes de familias reales a sintéticas, las lista de familias creadas se van a clasificar en 8 listas, una por cada tipo de familia, para así agrupar todas las listas de familias por tipos.

Funcionamiento del segundo submódulo

El preprocesamiento de los planes de la encuesta de movilidad es básico para la asignación correcta de los nuevos planes a la población sintética, éste es el segundo paso para la generación de los planes diarios de la población, la estructura de este submódulo se puede ver en la Figura 27.

El preprocesamiento fuera del programa permite arregla los problemas de la planificación diaria de algunas personas, agregandoles planes de vuelta a casa, eliminando hogares no aptos o modificando los destinos y orígenes de los viajes, permitiendo finalmente dividir los planes por tipologias. Este tercer submódulo obtiene uno de los ficheros csv con todos los planes de un tipo de familia, clasificando estos planes que va leyendo según la organización comentada en el apartado anterior.

En el quinto paso se selecciona un primer plan y se incluye dentro de la lista de planes de una persona, esto se continua haciendo para el resto de planes de esta persona hasta que se recibe un plan de una persona distinta pero que pertenece al mismo hogar, en ese caso crea una nueva lista de planes para la nueva persona, donde se incluirán todos sus planes. Este proceso continua con el resto de convivientes hasta que se recibe un plan de una persona nueva que no pertenece a la misma familia. Cuando esto pasa, se guarda la anterior familia en la lista del tipo de hogares a la que pertenecía esta casa, creando una nueva lista de familia y una nueva lista de planes de persona, donde se guardará el nuevo plan recibido.

Este proceso continúa hasta haber leído todos los planes del fichero csv, acabada la lectura de este fichero, se empieza con el siguiente, el cual contendrá los planes de un nuevo tipo de familia, planes que se procesarán de igual forma. Una vez preprocesados todos los tipos de familias, este tercer submódulo devuelve los planes organizados en 8 listas, una por cada tipo de hogar.

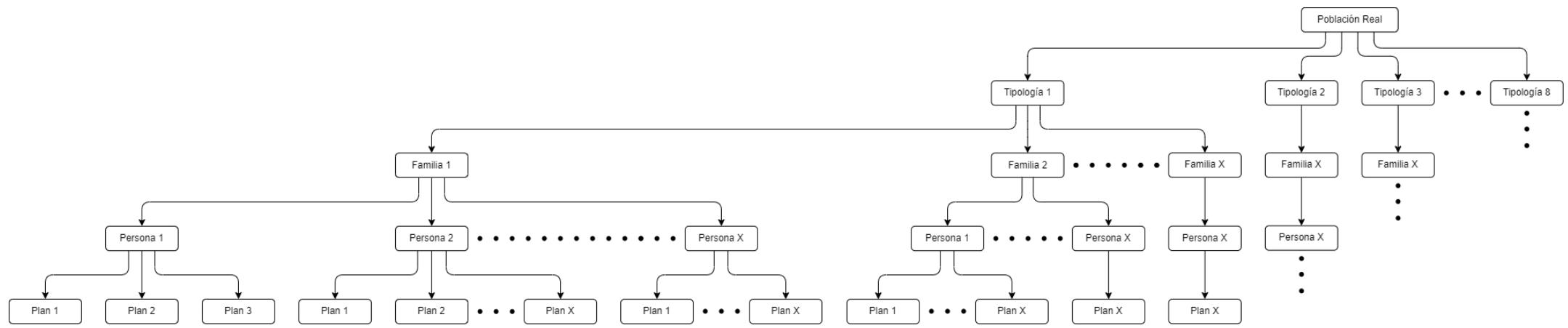


Figura 26. Diagrama organización de la población real

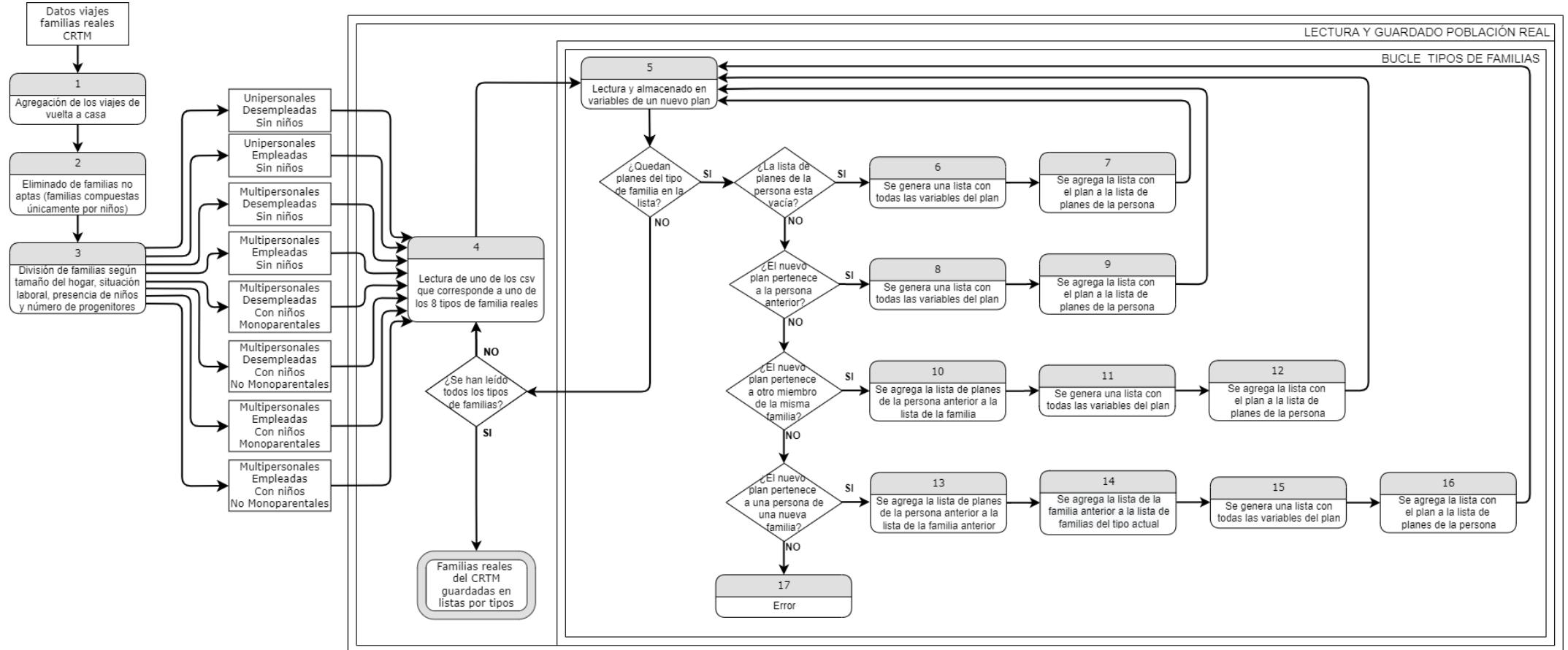


Figura 27. Diagrama submódulo 2

3.2.6.5. Traspaso de planes y escritura XML

Traspaso de planes

El traspaso de planes de una familia real a una familia sintética debe realizarse entre familias similares, por ello, tanto las familias sintéticas como las reales deben estar clasificadas por los mismo 8 tipos de familia. Cuando se seleccionada una familia sintética a completar con la planificación diaria, se debe seleccionar la lista que contiene los planes de las familias reales de la misma tipología. Además, es necesario comprobar que el número de personas que configuran ambas familias coincida, especialmente el número de adultos, que tienen unas planificación diarias más complejas.

Para cada familia sintética, se van a buscar hogares en la lista de familias reales hasta encontrar uno que coincida. Verificando para ello que ambos hogares contengan el mismo número de adultos, en caso contrario se descarta el hogar real, guardándolo al final de la lista y se pasa a comprobar el siguiente, este proceso de comprobaciones se va a realizar hasta que se encuentre una familia coincidente o hasta que se hayan comprobado todas las familias reales existentes sin encontrar una coincidencia, descartando en este caso la familia sintética.

Para incluir a familias jóvenes, dado que la edad elegida en el programa para diferenciar entre niños y adultos es 24 años, se comprueba que los hijos de la pareja joven sean lo suficientemente pequeños para que la madre haya dado a luz a partir de los 15. Cuando se da esta circunstancia, los jóvenes se contabilizan como adultos y se hace la comprobación del número de adultos entre ambas familias, si no fuera así, se descarta la familia sintética y se reinicia el proceso con una nueva.

Una vez comprobado el número de adultos, se revisarán los hijos. Si la familia real tiene más hijos que la familia sintética, entonces al disponer de planes reales de sobra, simplemente se borran los niños que no son necesarios. En caso contrario, si hay un niños reales de menos, cuando es solo uno, se duplica a uno de los hijos reales, pero en el caso de que haya dos o más hijos sintéticos de menos, la familia real se considera inválida y se pasará a comprobar con la siguiente familia de la lista.

Una vez seleccionada la familia real similar, se asigna para cada persona de la familia sintética una persona similar de la familia real. Para ello se ordenan por edad ambas familias, y para el caso de las parejas se comprueba el género. Si al ordenar las personas por edad no coincidiera el género de un miembro de la pareja real con el miembro de la pareja sintética que se le ha asignado, se modificará el orden de la familia real intercambiando de posición a los miembros de la pareja.

Las listas de familias reales contienen un número de familias mucho menor al número de personas sintéticas, por ello al ir sacando familias de la lista para realizar la transferencia de planes, estas listas se van vacían. Para evitar que la asignación de planes se quede sin familias reales de alguno de los tipos, estas lista son inicialmente copiadas en otras variables y en caso de vaciarse, se llenan con su copia original.

Escritura del fichero XML

El archivo XML en el que se escribe la población resultante sigue la sintaxis de los archivos de población de Matsim y consiste en una cabecera del archivo de población, escrita previa a este submódulo, seguida de una lista de personas escritas de forma independiente sin ser agrupadas por hogar o parentesco. Cada una de las personas contenida en este archivo tiene una cabecera con sus datos personales obtenidos de la familia sintética, seguida de la planificación diaria

ordenada por la hora de inicio de cada plan. La planificación incluye las distintas actividades que se realizan durante el día junto a los desplazamientos que se realizan para llegar a la siguiente actividad. Las actividades incluye datos como el tipo de actividad, las coordenadas donde se realiza la actividad o las horas de inicio y fin, por otro lado, los desplazamientos entre actividades contienen datos como el número de viaje, la duración de éste o el modo elegido para desplazarse, pudiendo optar entre el coche, el transporte público, el taxi, la bicicleta o andando.

Funcionamiento del tercer submódulo

La función principal del Planificador se realiza en este tercer submódulo, tras haber obtenido el ráster de Leganés del primer submódulo, los planes de las familias reales ya organizados del segundo submódulo y la población sintética entera del Familiador, este tercer submódulo realiza el propio traspaso de los planes diarios de unas familias a otras, haciendo uso del cuarto submódulo. La estructura de este submódulo se puede ver en la

Figura 28.

El proceso de traspaso de planes comienza con la selección de una familia sintética a completar con la planificación diaria, realizándose el proceso de búsqueda de una familia real acorde, tal y como se ha comentado anteriormente. Si se diera el caso de descartar esta familia sintética por no haber encontrado una familia real acorde, se comprobará que la lista de familias reales de ese tipo no contenga menos de la mitad de las familias originales, en caso afirmativo se rellena la lista de familias con la copia original de la misma lista. En caso de que coincidan los convivientes, se asignan a cada miembro de una familia los miembros correspondientes de la otra familia, teniendo en cuenta la edad y el género de las personas.

Una vez seleccionadas las dos familias con las que realizar el intercambio y fijados los miembros correspondientes de cada familia, se inicia el traspaso de planes y la escritura de la población sintética terminada en el XML que el programa proporcionará como *output*. Este proceso se va a llevar a cabo persona a persona, recogiendo los datos personales de la persona sintética, que se escriben como su cabecera en la salida, para a continuación ir obteniendo de uno en uno los datos de los planes de la persona real, modificando las ubicaciones por medio del cuarto submódulo que se explicará en el siguiente punto y escribiéndolos en el fichero de salida una vez terminados.

Finalmente, ya escritos todos los planes de la persona en el XML, se selecciona al siguiente conviviente si es que lo hubiera, realizando este mismo proceso, el cual termina al haber escrito a todas las personas de la familia, momento en el cual, se reiniciará el proceso para completar la siguiente familia sintética.

3.2.6.6. Obtención de nuevas coordenadas

El cuarto y último submódulo del Planificador se encarga de generar las coordenadas que indican el lugar desde el que se realizan las actividades de la planificación diaria, adaptando los planes de la encuesta de movilidad a cada persona, este submódulo es llamado por el tercer submódulo cada vez que se necesite calcular un nuevo destino.

El cálculo de las nuevas coordenadas difiere según el desplazamiento que se tenga realizar. Se distinguen tres tipos de desplazamientos, uno primero que agrupa los desplazamientos al domicilio, un segundo tipo de desplazamiento para las personas que salen del municipio de Leganés y un tercer tipo de desplazamiento para los viajes que se producen dentro del municipio de Leganés cuando el destino no es el domicilio.

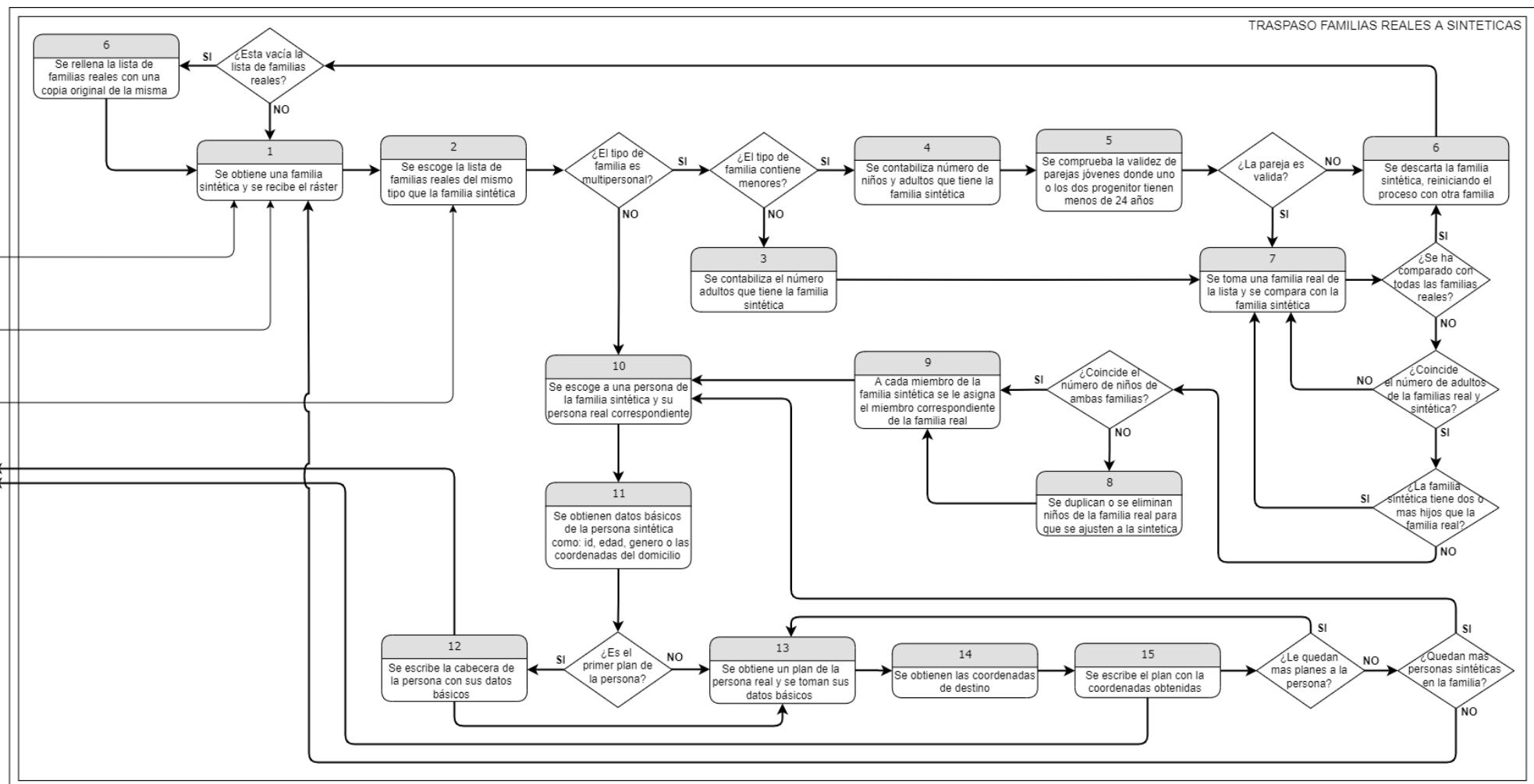


Figura 28. Diagrama submódulo 3

Desplazamiento al domicilio

En el caso de volver a la residencia no se tiene que realizar ningún cálculo, ya que se tienen las coordenadas exactas del domicilio real obtenido del Catastro, que le fue asignado a cada persona en el Familiador.

Desplazamiento fuera del municipio

Para los viajes que salían de Leganés, debido a que estas personas salían del ráster creado, no se podría calcular las coordenadas de destino, ni la ruta seguida para llegar a los nuevos municipios, por otro lado, la movilidad que se quería simular era la que interviene dentro del municipio, con lo que no era necesario calcular los destinos con exactitud. Sin embargo, los tramos iniciales hasta salir del municipio o los tramos finales de llegada de estos viajes se producían dentro del ráster de Leganés y por ello afectaban a la movilidad dentro del municipio, como este tipo de desplazamientos abundaban en la encuesta del CRTM, merecían ser incluidos.

Los desplazamientos fuera del municipio eran muy frecuentes debido a que Leganés es una “ciudad dormitorio”, lo que supone que gran cantidad de los vecinos del municipio trabajan fuera de éste. Como consecuencia los días laborables las vías de salida del municipio tienden a tener un aumento del tráfico, produciéndose retenciones en algunos casos.

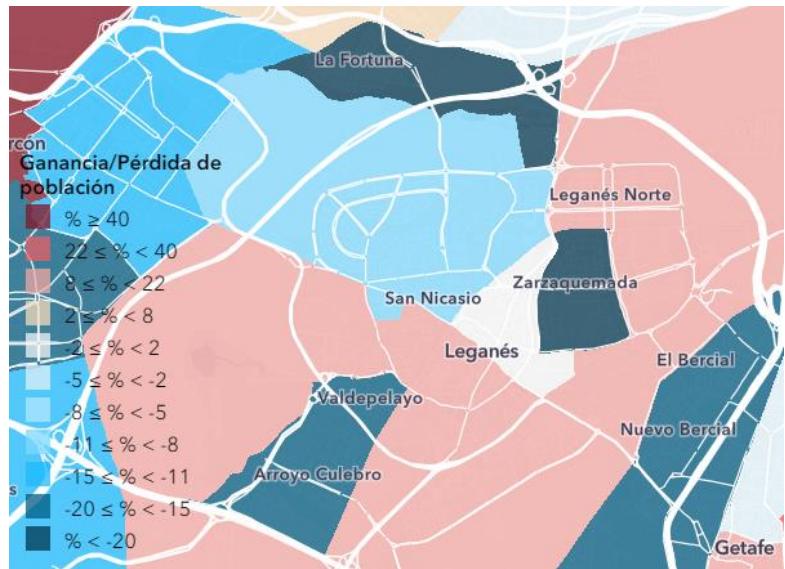


Figura 29. Mapa de desplazamientos INE [60]

Esto se puede ver gracias al estudio de movilidad por telefonía móvil realizado por el INE [60] que incluía un mapa de España dividido por distritos, resaltando la pérdida o ganancia de población que se muestra en la Figura 29.

Este mapa del INE muestra como ciertas partes de Leganés, especialmente el barrio de Zarzaquemada y el de La Fortuna pierden población en horario laboral, que se desplaza fuera del barrio para trabajar, este hecho se invierte en otras zonas que reciben mayor población de la que sale, aunque este fenómeno se da en menores porcentajes.

Como no se podía calcular las coordenadas de destino exactas de estos viajes, pero sí se conocía el municipio de destino, se buscaron las vías de comunicación que conectaban Leganés con las ciudades destino. Tras una búsqueda de las vías de salida más transitadas, se obtuvieron los puntos de salida del municipio que más convenían a los leganenses para realizar estos trayectos. Como el número de poblaciones aparecidas en la encuesta del consorcio de transportes era muy alto, únicamente se seleccionaron las ciudades con las que hubiera un mayor número de desplazamientos, sustituyendo el resto de las ciudades por una de las seleccionadas.

En la **Figura 30** se muestran los puntos de salida de Leganés por los que deberían salir las personas que deseen ir a uno de los municipios seleccionados, que se muestran en su margen.

Para calcular las coordenadas destino de este tipo de viajes, el submódulo simplemente devolverá las coordenadas de salida de Leganés asociadas al municipio de destino, que ya tiene guardadas, por lo que no se tiene que realizar ningún cálculo.



Figura 30. Puntos de salida del municipio de Leganés

Desplazamiento dentro del municipio

Cuando la residencia no es el destino del viaje, no se tienen las coordenadas destino y por ello hay que calcular unas nuevas. Este tipo de desplazamiento es el más complejo de todos, pues se va a tener que calcular las nuevas coordenadas destino, haciendo uso de las coordenadas de origen, de la distancia recorrida, del tipo de actividad y del ráster generado en el primer submódulo, que contiene la información de los inmuebles de Leganés.

El primer paso para calcular las nuevas coordenadas de este tipo de plan pasa por aplicar la distancia recorrida durante el viaje sobre el punto origen, es decir sobre la ubicación del plan anterior. Si se traza una circunferencia sobre este punto, de radio la distancia recorrida, se pueden identificar todos los puntos posibles que se podrán escoger como destino.

En la Figura 31 se ha trazado una circunferencia cuyo radio es la distancia recorrida sobre el ráster de Leganés. Como ya se comentó, el cálculo del destino depende del conjunto de edificios que se encuentran alrededor del nuevo punto hallado, es decir, para elegir un punto de destino se utilizan los cuadrantes de los que se compone el ráster y para ello, habrá que identificar las celdas por las que pasa la circunferencia.



Figura 31. Cálculo del nuevo destino 1

Para conseguir hallar esta celdas, se van a trazar puntos alrededor de la circunferencia, para posteriormente identificar la celda en la que se encuentran cada uno de los puntos generados. Sin embargo, la cantidad de puntos a calcular para obtener las celdas va a ser mayor o menor dependiendo de las distancias, en la Figura 33, se puede apreciar que el número de puntos calculado es innecesario. Por ello para cada distancia se van a calcular una cantidad específica de puntos como se puede ver en la última figura, calculando puntos cada cierto número de grados girados, para distancias de menos de 500 metros, se generan puntos cada 20° , para distancias de 500m a 1000m cada 10° , para distancias de 1000m a 3000m cada 5° y, por último, para distancias de más de 3000m se calcularán puntos cada 2.5° .

Como puede darse el caso de que, al aplicar una distancia grande sobre un punto situado en el centro de Leganés, la circunferencia quede fuera del ráster, en los casos en que esto ocurra, se limita la distancia, creando una circunferencia de menor radio que interseque con distintas celdas para calcular el destino.

El cálculo del primer punto es el más sencillo, para obtenerlo solo se le tienen que sumar la distancia recorrida en el eje X a las coordenadas de origen. Para calcular los siguientes puntos se utilizarán las razones trigonométricas, pasando los grados a radianes y obteniendo el seno y el coseno del ángulo de giro, con esta operación se obtiene el siguiente punto. Aprovechando la simetría de la circunferencia, solo se tienen que calcular mediante este método los puntos del primer cuadrante. Calculando los puntos del segundo cuadrante a partir de los del primer cuadrante utilizando la simetría del eje y, y calculando los puntos del tercer y cuarto cuadrante a partir de los del primer y segundo cuadrante utilizando con el eje x como eje de simetría.

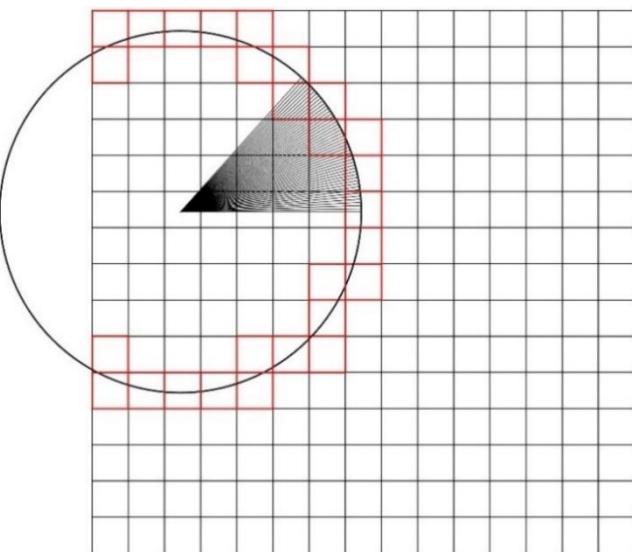


Figura 32. Cálculo del nuevo destino 2

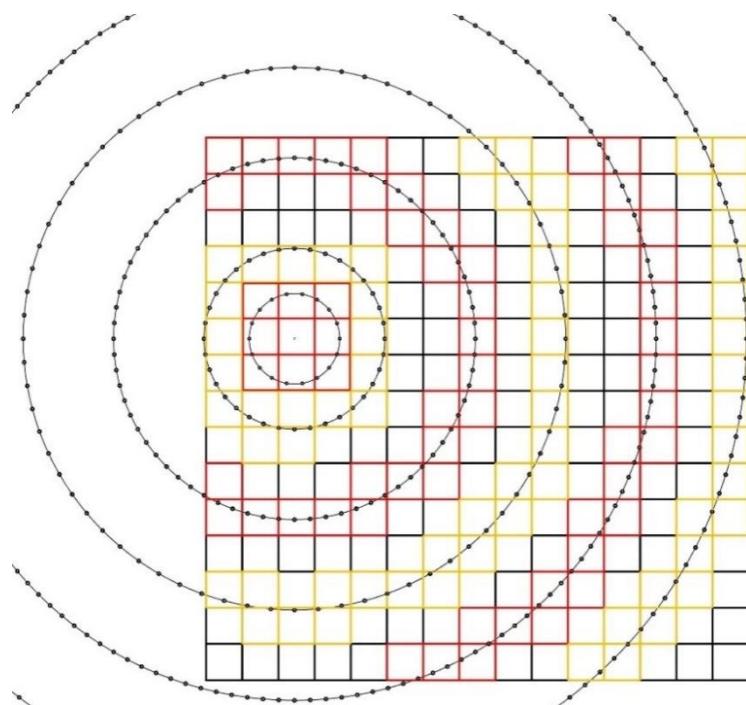


Figura 33. Cálculos del nuevo destino 3

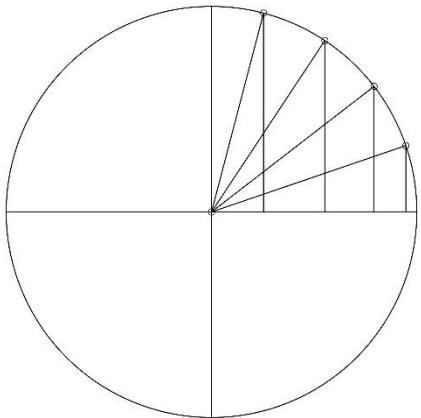


Figura 34. Generación de nuevas coordenadas 1

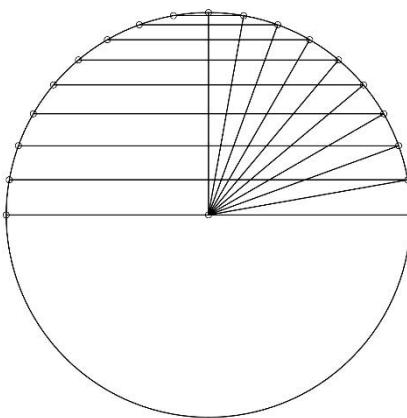


Figura 35. Generación de nuevas coordenadas 2

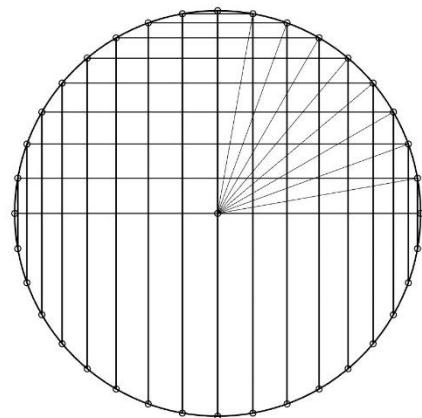


Figura 36. Generación de nuevas coordenadas 3

Una vez obtenidos los puntos, se calculan las celdas en las que se encuentran estos, guardando las celdas halladas en una lista junto a las coordenadas del punto con las que se hallaron. Como se puede dar el caso de que dos o más puntos coincidan en una misma celda, cuando esto ocurra, la celda solo se introducirá una vez en la lista y se descartarán el resto de los puntos. Para cada celda hallada se va a obtener la puntuación del tipo de actividad asociada al plan, guardando las puntuaciones con sus celdas en la lista.

Por último, para elegir el destino final se va a escoger una de las celdas de la lista valorando las puntuaciones de cada una de ellas como probabilidades de ser elegidas, mediante un número aleatorio se elegirá una de las celdas, obteniendo las coordenadas destino asociadas a esta, mediante las cuales se había identificado previamente la celda.

Funcionamiento del cuarto submódulo

Este cuarto submódulo del Planificador asiste la generación de los planes del tercer submódulo, generando las coordenadas destino de tres formas distintas, en función del tipo de desplazamiento que se requiera. La estructura de este submódulo se puede ver en el diagrama de la Figura 37.

El proceso de generación de planes se inicia cuando el tercer submódulo envía un plan al generador de coordenadas. El plan contendrá las coordenadas origen desde donde se iniciará el desplazamiento, el tipo de actividad que se va a realizar en el destino, la distancia entre origen y destino y el código de municipio en el que se realiza la actividad.

Lo primero que hace este submódulo es averiguar a cuál de los tres tipos de desplazamientos pertenece el plan original. En caso de que el viaje tenga como tipo de actividad “*home*”, se devolverán las coordenadas de la casa a la que pertenece esta persona, en el caso de que el código de municipio destino no sea el código de Leganés, se devolverán las coordenadas del punto de salida de Leganés asignadas al municipio del código. Por último, en caso de que el plan original no sea de los dos tipos anteriores, significará que el desplazamiento se realiza dentro del municipio y que el tipo de actividad no es “*home*”, calculando mediante el método explicado en el apartado anterior las nuevas coordenadas destino.

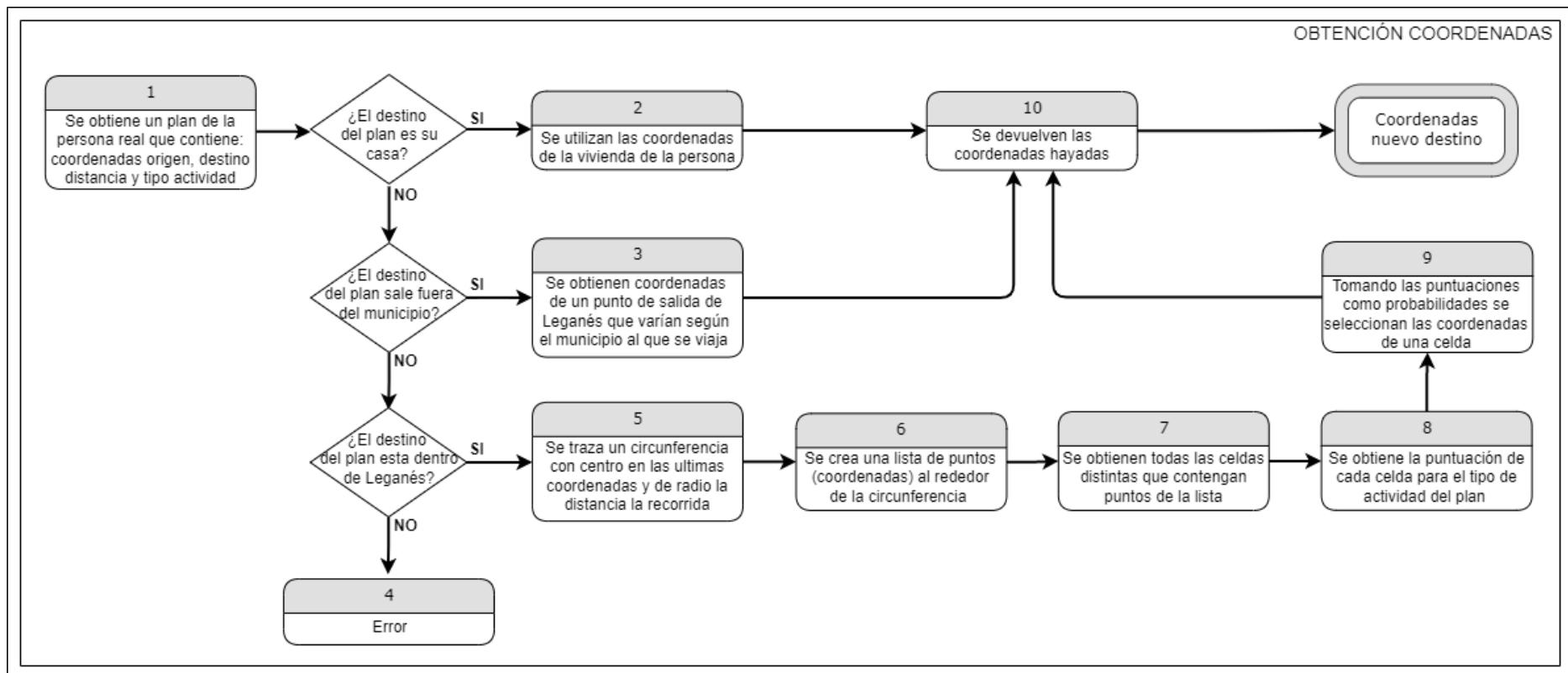


Figura 37. Diagrama submódulo 4

3.2.6.7. Mejoras

Al igual que con el Familiador, este segundo módulo de The Synthetic Population, admite una serie de mejoras que no pudieron ser implementadas en la primera versión. En este apartado se van a enumerar distintos puntos de mejora para indicar posibles caminos de desarrollo a investigadores interesados en el proyecto y así facilitar el desarrollo de este software.

Creación de los planes diarios en base a un análisis sociológico

Para modificar esta sección del programa habría que estudiar la viabilidad de este campo, se tendrían que detectar los tipos de actividades más realizadas, los horarios en que se realizan las actividades, la duración de estas, diferenciar las actividades según la tipología de familia, etc.

La aplicación de este modelo no es una tarea fácil, sin embargo, continuar con el enfoque sociológico del Familiador en el Planificador facilitaría la asignación de planes a las familias sintéticas, pues la asignación de planes a las familias sintéticas ya no dependería en su totalidad de los viajes reales en bruto de una encuesta de movilidad, cuya escasez de datos hizo difícil su clasificación, obligando a la creación de 8 tipologías más generales para conectar el Familiador con el Planificador. La asignación de planes dependería de las estadísticas extraídas de los viajes (tipo, hora, edad, etc.) y permitirían generar los planes junto a la creación de las familias dentro de las tipologías específicas del Familiador, obviando crear las 8 tipologías más generales que poseen una información más reducida de los hogares.

Mejora de la tipologías implementadas

En el supuesto de no aplicar la generación de planes en base a la sociología, tal y como se indicaba en el punto anterior, se podría mejorar la subdivisión actual basada en 8 tipos de hogares, agregando nuevos factores que generen más tipos de hogares, con el fin de aproximarse a las tipologías desarrolladas en el Familiador, consiguiendo un movimiento más realista. Esta mejora, como ya se comentó anteriormente, está condicionada por los datos disponibles en la encuesta de movilidad utilizada en el proyecto, que no permitió una especificación mayor en las clases.

Factores

Un factor destacable que se debería introducir en nuevas versiones del proyecto sería el teletrabajo y es que durante la pandemia ha aumentado considerablemente la cantidad de personas que trabajan desde casa. El hecho de que las estadísticas utilizadas fueran previas a la pandemia causada por el Covid, ha impedido que este factor fuera introducido en la primera versión, pues en 2019 el teletrabajo no influía significativamente en la movilidad. La versión actual del programa no hace una distinción entre trabajadores. Incluyendo indistintamente los planes de personas que se desplazan al lugar de trabajo y los de las personas que teletrabajan a la población que dispone de un empleo. Versiones futuras del trabajo podrían incluir este factor en el software haciendo una estimación de la cantidad de trabajadores que operan desde casa en función del sector profesional, atribuyendo porcentajes de incidencia del teletrabajo para cada sector.

Vinculación de familias reales y sintéticas

La asignación persona a persona de las familias reales a las sintéticas, a excepción de los hogares unipersonales, se basa en la edad y en el género en el caso de las parejas. Por ello, sería recomendable incluir en futuras versiones, como mínimo, estos mismos parámetros para los hogares unipersonales, aunque también sería recomendable incorporar otros parámetros diferenciadores, como las relaciones de parentesco entre los convivientes. Esta mejora lograría establecer una asignación más compleja de los miembros de los hogares, consiguiendo que las planificaciones diarias se ajusten mejor a la realidad.

Ráster

El sistema ideado para generar las coordenadas de salida del ráster de Leganés incluye unas coordenadas específicas que no varían, aunque las personas que se desplacen al nuevo municipio se encuentren en distintas ubicaciones dentro de Leganés. En el caso de darse desplazamientos a municipios lejanos a Leganés, tener un solo punto de salida no supone un problema. Sin embargo, para desplazamientos a municipios cercanos, como podría ser el caso de Alcorcón, el trayecto se podría realizar por cualquiera de las múltiples vías que comunican ambas ciudades, siendo las personas las que deciden que camino tomar, soliendo considerar su ubicación actual y la ubicación de destino dentro de Alcorcón. Por ello, sería conveniente generar distintas coordenadas de salida de Leganés para cada municipio destino según la ubicación origen de las personas, para así seleccionar el trayecto con la menor duración posible al punto de destino, lo que supondría una mejora considerable.

Otro punto que mejorar del ráster radica en la adquisición de datos más precisos. Como ya se comentó, los datos ofrecidos por el catastro no contenían la suficiente información de los inmuebles como para poder realizar una mejor asignación de los destinos. La sustitución/combinación de los datos del Catastro con los datos de otras fuentes como OSM podría ser una solución a este problema.

Población

Ciertos problemas que se arrastran del Familiador continúan en el Planificador. Este es el caso de la situación laboral, que se asigna al hogar y no a las personas. Esto implica que, en la asignación de la planificación diaria, personas que en realidad están empleadas se procesen como desempleadas, introduciendo una movilidad no adecuada a estas personas.

La población real con la que se trabaja en el Planificador, escogida de la encuesta del CRTM, únicamente tiene su residencia en Leganés¹¹, esto significa que se está trabajando con la población saliente de Leganés, pero no con la población entrante. Para mejorar el estudio de circulación de Leganés no se debería incluir únicamente a la población que reside en Leganés, también se debería incluir a todas aquellas personas que tienen su residencia en otros municipios, pero que se desplazan diariamente a Leganés para trabajar o para realizar otro tipo de actividades. Por poner un ejemplo, en el caso de que una persona tenga su domicilio en Fuenlabrada, pero trabaje en Leganés, su movilidad no queda recogida en la versión actual del trabajo. Como es más difícil establecer su domicilio o los planes que realiza esta persona dentro de su municipio de origen y además estos desplazamientos no son el objetivo del estudio, se podría trabajar únicamente con los planes que realizan dentro del municipio de Leganés.

¹¹ La población obtenida de Getafe, Alcorcón y Móstoles se van a tratar como empadronados en Leganés

4. Análisis y evaluación del sistema

En este apartado se van a detallar las necesidades y el comportamiento que debe tener el software que se desarrolla, mediante la especificación de los casos de uso y de los requisitos de software, que deberá cumplir el sistema.

4.1. Casos de uso

4.1.1. Tabla de requisitos y nomenclatura

La plantilla que se va a utilizar para especificar los casos de uso tendrá el siguiente formato:

CU-XX	
Nombre	Nombre del caso de uso
Descripción	Descripción del caso de uso
Postcondiciones	Condiciones que se dan tras el caso de uso
Condiciones de error	Condiciones que podrían causar un error

Tabla 6. Plantilla tabla casos de uso

La tabla mostrada arriba está formada por los siguientes componentes:

- **Identificador:** Compuesto de la siguiente forma “CU-XX”, donde CU hace referencia al caso de uso y XX al número de caso de uso.
- **Descripción:** Breve definición del caso de uso.
- **Postcondiciones:** Estado del sistema tras la ejecución de la operación
- **Condiciones de error:** Condiciones que causarían que durante la ejecución se diera un error en el sistema

4.1.2. Especificación de Casos de Uso

CU-01	
Nombre	Ajuste de estadísticas
Descripción	El usuario debe intercambiar todos los porcentajes de selección del municipio de Leganés por los de la región en estudio
Postcondiciones	El programa contiene las estadísticas
Condiciones de error	Las cifras introducidas no tienen un correcto formato

Tabla 7. Caso de uso CU-01

CU-02	
Nombre	Preparación previa
Descripción	El usuario debe almacenar en el directorio del ejecutable los ficheros csv del catastro, del censo y de la encuesta de viajes
Postcondiciones	El programa está listo para ejecutar
Condiciones de error	-

Tabla 8. Caso de uso CU-02

CU-03	
Nombre	Ejecución
Descripción	Se ejecuta el programa
Postcondiciones	Se ha obtenido un fichero XML con la población
Condiciones de error	Los ficheros csv no se encuentran en el directorio del ejecutable o su nombre o formato son incorrectos

Tabla 9. Caso de uso CU-03

4.2 Requisitos de software

4.2.1. Tabla de requisitos y nomenclatura

La plantilla que se va a utilizar para especificar los requisitos de software tendrá el siguiente formato:

RS-XX-YY			
Nombre	Nombre del requisito de software		
Descripción	Descripción del requisito de software		
Prioridad	Alta	Media	Baja
Necesidad	Esencial	Deseable	Opcional
Verificabilidad	Alta	Media	Baja
Estabilidad	Si		No

Tabla 10. Plantilla tabla requisitos de software

La tabla mostrada arriba está formada de los siguientes componentes:

- **Identificador:** Compuesto de la siguiente forma “RS-XX-YY”, donde RS hace referencia a requisito de software, YY al número de requisito de software y XX al tipo de requisito seleccionado de entre los siguientes:
 - o RF: Requisito funcional
 - o NF: Requisito no funcional
- **Descripción:** Breve definición del requisito.
- **Prioridad:** Grado de importancia que planifica el orden en el que se implementaran las distintas características enunciadas en los requisitos, con tres niveles de prioridad:
 - o Alta
 - o Media
 - o Baja
- **Necesidad:** Grado de importancia de un requisito software, con tres niveles de necesidad:
 - o Esencial
 - o Deseable
 - o Opcional
- **Verificabilidad:** Grado de facilidad con la que se pueden comprobar la correcta incorporación del requisito al software, con tres niveles de verificabilidad:
 - o Alta
 - o Media
 - o Baja
- **Estabilidad:** Indica si un requisito es susceptible a modificaciones.

4.2.2 Especificación de Requisitos funcionales

RS-RF-01			
Nombre	<i>Input</i> en el sistema		
Descripción	El software estará preparado para leer datos de archivos con extensión csv		
Prioridad	Alta	Media	Baja
Necesidad	Esencial	Deseable	Opcional
Verificabilidad	Alta	Media	Baja
Estabilidad	Si	No	

Tabla 11. Requisito de software RS-RF-01

RS-RF-02

Nombre	Nombre del requisito de software		
Descripción	El sistema generará familias sintéticas a partir de estadísticas de la población real objetivo		
Prioridad	Alta	Media	Baja
Necesidad	Esencial	Deseable	Opcional
Verificabilidad	Alta	Media	Baja
Estabilidad	Si	No	

Tabla 12. Requisito de software RS-RF-02

RS-RF-03			
Nombre	Nombre del requisito de software		
Descripción	El sistema generará números aleatorios para elegir entre las distintas tipologías de familia existentes		
Prioridad	Alta	Media	Baja
Necesidad	Esencial	Deseable	Opcional
Verificabilidad	Alta	Media	Baja
Estabilidad	Si	No	

Tabla 13. Requisito de software RS-RF-03

RS-RF-04			
Nombre	Nombre del requisito de software		
Descripción	Seleccionada una tipología de hogar, el sistema utilizará números aleatorios para seleccionar el sexo y la edad de las personas que conforman el hogar		
Prioridad	Alta	Media	Baja
Necesidad	Esencial	Deseable	Opcional
Verificabilidad	Alta	Media	Baja
Estabilidad	Si	No	

Tabla 14. Requisito de software RS-RF-04

RS-RF-05			
Nombre	Nombre del requisito de software		

Descripción	El sistema seleccionará personas del censo de Leganés para generar las familias		
Prioridad	Alta	Media	Baja
Necesidad	Esencial	Deseable	Opcional
Verificabilidad	Alta	Media	Baja
Estabilidad	Si		No

Tabla 15. Requisito de software RS-RF-05

RS-RF-06			
Nombre	Nombre del requisito de software		
Descripción	El sistema generará un objeto de familia con los datos de las personas seleccionadas, que será almacenado en una lista de objetos		
Prioridad	Alta	Media	Baja
Necesidad	Esencial	Deseable	Opcional
Verificabilidad	Alta	Media	Baja
Estabilidad	Si		No

Tabla 16. Requisito de software RS-RF-06

RS-RF-07			
Nombre	Nombre del requisito de software		
Descripción	El sistema asignara a cada familia un identificador unico		
Prioridad	Alta	Media	Baja
Necesidad	Esencial	Deseable	Opcional
Verificabilidad	Alta	Media	Baja
Estabilidad	Si		No

Tabla 17. Requisito de software RS-RF-07

RS-RF-08			
Nombre	Nombre del requisito de software		
Descripción	El sistema asignará una casa real a cada familia teniendo en cuenta el número de integrantes de los que se compone		

Prioridad	Alta	Media	Baja
Necesidad	Esencial	Deseable	Opcional
Verificabilidad	Alta	Media	Baja
Estabilidad	Si		No

Tabla 18. Requisito de software RS-RF-08

RS-RF-09			
Nombre	Nombre del requisito de software		
Descripción	El sistema asignará a cada hogar una de las 8 tipologías de familia existentes, teniendo en cuenta el número de integrantes, la situación laboral, la presencia de menores en el hogar y la presencia de un solo adulto en el caso de haber menores		
Prioridad	Alta	Media	Baja
Necesidad	Esencial	Deseable	Opcional
Verificabilidad	Alta	Media	Baja
Estabilidad	Si		No

Tabla 19. Requisito de software RS-RF-09

RS-RF-10			
Nombre	Nombre del requisito de software		
Descripción	El sistema generará un conjunto de familias para cada distrito		
Prioridad	Alta	Media	Baja
Necesidad	Esencial	Deseable	Opcional
Verificabilidad	Alta	Media	Baja
Estabilidad	Si		No

Tabla 20. Requisito de software RS-RF-10

RS-RF-11			
Nombre	Nombre del requisito de software		
Descripción	El sistema generará un ráster de Leganés compuesto por objetos de celda		
Prioridad	Alta	Media	Baja

Necesidad	Esencial	Deseable	Opcional
Verificabilidad	Alta	Media	Baja
Estabilidad	Si		No

Tabla 21. Requisito de software RS-RF-11

RS-RF-12			
Nombre	Nombre del requisito de software		
Descripción	Cada celda del ráster contendrá 4 coordenadas que definen el área exacta que encierran dentro.		
Prioridad	Alta	Media	Baja
Necesidad	Esencial	Deseable	Opcional
Verificabilidad	Alta	Media	Baja
Estabilidad	Si		No

Tabla 22. Requisito de software RS-RF-12

RS-RF-13			
Nombre	Nombre del requisito de software		
Descripción	Cada celda del ráster contendrá 5 variables correspondientes a cada uno de los tipos de actividades existentes: work, medical, leisure, shopping, education.		
Prioridad	Alta	Media	Baja
Necesidad	Esencial	Deseable	Opcional
Verificabilidad	Alta	Media	Baja
Estabilidad	Si		No

Tabla 23. Requisito de software RS-RF-13

RS-RF-14			
Nombre	Nombre del requisito de software		
Descripción	El sistema leerá una lista de los edificios de Leganés con sus coordenadas y sus usos.		
Prioridad	Alta	Media	Baja
Necesidad	Esencial	Deseable	Opcional

Verificabilidad	Alta	Media	Baja
Estabilidad	Si		No

Tabla 24. Requisito de software RS-RF-14

RS-RF-15			
Nombre	Nombre del requisito de software		
Descripción	El sistema asignará a cada celda del ráster una calificación por cada uno de los tipos de actividades existentes usando la lista de edificios del Catastro		
Prioridad	Alta	Media	Baja
Necesidad	Esencial	Deseable	Opcional
Verificabilidad	Alta	Media	Baja
Estabilidad	Si		No

Tabla 25. Requisito de software RS-RF-15

RS-RF-16			
Nombre	Nombre del requisito de software		
Descripción	El sistema traspasara los viajes de una familia real a una familia sintética de la misma tipología		
Prioridad	Alta	Media	Baja
Necesidad	Esencial	Deseable	Opcional
Verificabilidad	Alta	Media	Baja
Estabilidad	Si		No

Tabla 26. Requisito de software RS-RF-16

RS-RF-17			
Nombre	Nombre del requisito de software		
Descripción	El programa asignará coordenadas de destino predefinidas a los viajes que se realizan fuera del Municipio de Leganés		
Prioridad	Alta	Media	Baja
Necesidad	Esencial	Deseable	Opcional
Verificabilidad	Alta	Media	Baja

Estabilidad	Si	No
-------------	----	----

Tabla 27. Requisito de software RS-RF-17

RS-RF-18			
Nombre	Nombre del requisito de software		
Descripción	El programa genera coordenadas de destino a los viajes que se realizan dentro de Leganés a partir de la distancia del desplazamiento y de la calificación de las actividades del ráster que usará como pesos		
Prioridad	Alta	Media	Baja
Necesidad	Esencial	Deseable	Opcional
Verificabilidad	Alta	Media	Baja
Estabilidad	Si	No	

Tabla 28. Requisito de software RS-RF-18

RS-RF-19			
Nombre	Nombre del requisito de software		
Descripción	El sistema generará un archivo con extensión XML que contendrá los desplazamientos de cada persona		
Prioridad	Alta	Media	Baja
Necesidad	Esencial	Deseable	Opcional
Verificabilidad	Alta	Media	Baja
Estabilidad	Si	No	

Tabla 29. Requisito de software RS-RF-19

4.2.3. Especificación de Requisitos no funcionales

RS-NF-01			
Nombre	Lenguaje de programación		
Descripción	El software deberá ser programado en Python		
Prioridad	Alta	Media	Baja
Necesidad	Esencial	Deseable	Opcional
Verificabilidad	Alta	Media	Baja

Estabilidad	Si	No
-------------	----	----

Tabla 30. Requisito de software RS-NF-01

RS-NF-02			
Nombre	Nombre del requisito de software		
Descripción	El sistema deberá informar al usuario en el caso de que al leer un fichero con datos esté vacío		
Prioridad	Alta	Media	Baja
Necesidad	Esencial	Deseable	Opcional
Verificabilidad	Alta	Media	Baja
Estabilidad	Si	No	

Tabla 31. Requisito de software RS-NF-02

RS-NF-03			
Nombre	Nombre del requisito de software		
Descripción	El sistema almacenará los datos leídos en listas para su posterior uso		
Prioridad	Alta	Media	Baja
Necesidad	Esencial	Deseable	Opcional
Verificabilidad	Alta	Media	Baja
Estabilidad	Si	No	

Tabla 32. Requisito de software RS-NF-03

RS-NF-04			
Nombre	Nombre del requisito de software		
Descripción	Deberá proporcionarse un manual de usuario junto al software desarrollado.		
Prioridad	Alta	Media	Baja
Necesidad	Esencial	Deseable	Opcional
Verificabilidad	Alta	Media	Baja
Estabilidad	Si	No	

Tabla 33. Requisito de software RS-NF-04

RS-NF-05			
Nombre	Nombre del requisito de software		
Descripción	El fichero XML generado como salida deberá seguir el formato específico definido por MatSim		
Prioridad	Alta	Media	Baja
Necesidad	Esencial	Deseable	Opcional
Verificabilidad	Alta	Media	Baja
Estabilidad	Si	No	

Tabla 34. Requisito de software RS-NF-05

4.3 Evaluación

Una vez enumerados los casos de uso y los requisitos del programa, se va a comprobar si la población sintética resultante del software desarrollado en este trabajo se ajusta a la población real, demostrando así la validez del programa. Para ello se van a comparar las 8 tipologías de hogares obtenidas de *The Synthetic Population*, con las mismas 8 tipologías obtenidas de los microdatos del INE, comprobando si los porcentajes de uno se corresponden con los del otro.

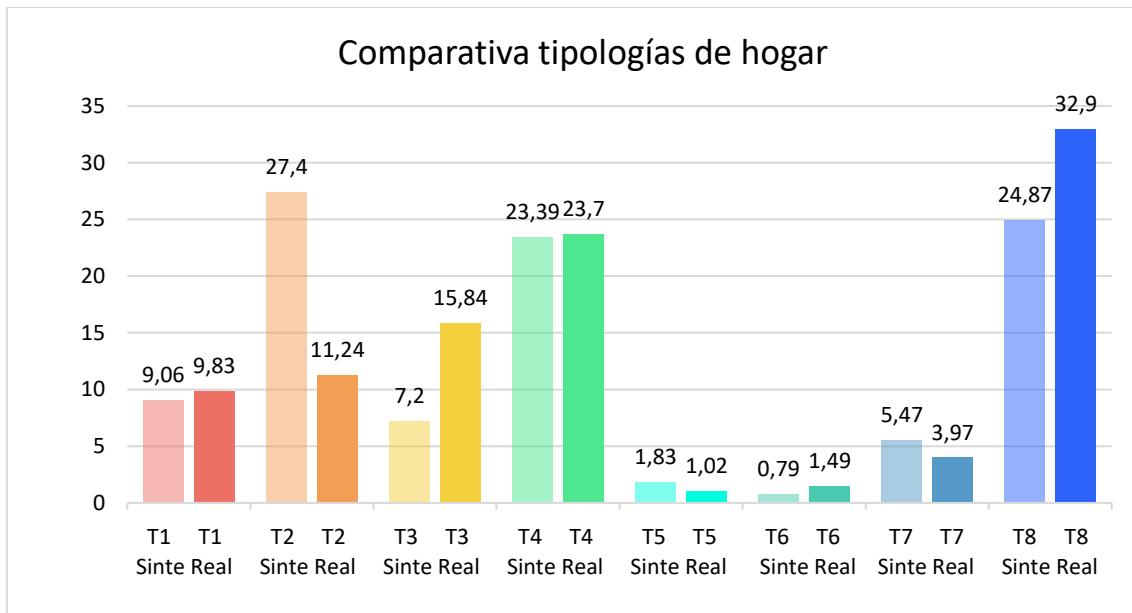


Figura 38. Comparativa entre tipologías de hogar reales y sintéticas

Como se puede ver en la Figura 38, algunas tipologías de hogares se ajustan mejor que otras, el caso que menos se ajusta de la comparativa corresponde a los hogares unipersonales con empleo (Tipo 2), para esta tipología el Familiador ha calculado que representan un 27,4% del total de hogares es de este tipo, frente al porcentaje real de INE del 11,24%. Por otro lado, los hogares multipersonales en paro y sin hijos (Tipo 3) están infrarrepresentados por el Familiador que les atribuye un 7,2% de la incidencia cuando debieran representar un 15,84%.

Estos problemas de ajuste de deben en parte a que el primer prototipo del software desarrollado no contiene la mayoría de las tipologías de hogar identificadas. Tal y como ya se ha explicado previamente, la cantidad de tiempo requerido para el desarrollo de esta área habría limitado el objetivo principal de este trabajo, por lo que se limitó la generación de todas las tipologías. Con el futuro desarrollo del programa se lograría una mejora de los resultados obtenidos, consiguiendo que los hogares sintéticos generados se ajustaran en mayor medida a los hogares reales.

Otra comparación que se hace necesaria es la de los planes de la población sintética, con los planes obtenidos de la encuesta de movilidad del Consorcio de transportes. Para ello, se han equiparado porcentualmente el número de actividades de cada tipo desarrolladas durante la planificación diaria, obteniendo los resultados que se muestran en las Figuras 39 y 40.

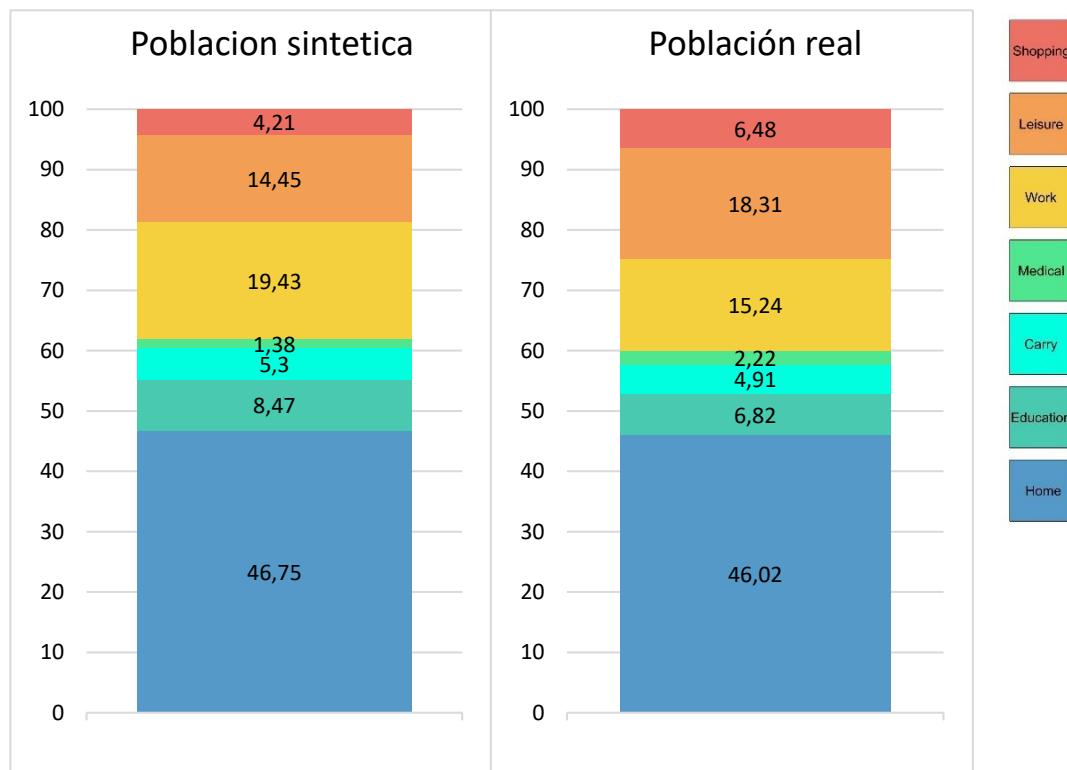


Figura 39. Actividades de la población sintética

Figura 40. Actividades de la población real

Los resultados obtenidos en esta segunda comparación son mejores que los de la primera. Los tipos de actividades realizadas por la población sintética se ajustan en mayor medida a los de la población real. Sin embargo, este ajuste no es perfecto y esto en parte es debido a que el mal ajuste de las tipologías de hogares observado en la gráfica de la Figura 38. La infrarrepresentación o sobrerrepresentación de algunos tipos de hogares, conlleva la infrarrepresentación o sobrerrepresentación de algunos tipos de actividades, esto es debido a que algunos tipos de hogares realizan mayoritariamente unos tipos de actividades que otras tipologías no realizan. Por poner ejemplo, aparece una mayor cantidad de desplazamientos de tipo *educational* como consecuencia del mayor número de familias simuladas con niños. Esto es debido a que muchas tipologías de familias que contienen un mayor número de hijos no fueron implementadas en el Familiador, por ello, los niños del censo de Leganés se repartieron entre más familias con un solo hijo, lo que ha llevado a un aumento del tipo de actividad *educational*.

5. Simulación

La simulación en Matsim requiere como mínimo tres archivos, el archivo *config*, el *network* y el *population*, también denominado *plans* en algunos proyectos.

- **Config.xml:** Este archivo sirve para configurar las simulaciones, contiene para ello una lista de parámetros y los valores de estos. Los parámetros están agrupados en módulos según la función que desarrollan, uno de ellos engloba la variables para puntuar los planes, e incluye todos los tipos de actividades que se pueden realizar y las recompensas que se atribuyen cada vez que ocurren una de las situaciones definidas, como llegar tarde al destino. Otro módulo importante es el de la estrategia, el cual especifica las estrategias para replanificar los desplazamientos, pudiendo cambiar las horas a las que se realizan estos planes, el tipo de transporte elegido o incluso el plan escogido por otro, en el caso de tener más de uno. Un último módulo a resaltar es el de control, el cual especifica las iteraciones que se van a realizar o la dirección del output. Matsim ofrece una gran cantidad de módulos distintos que pueden ser configurados por el usuario
- **Network.xml:** Este archivo contiene la red viaria sobre la que se va a simular y debe ser referenciado en el archivo *config* para poder realizar la simulación. La representación de la red viaria se hace mediante un grafo, donde las calles son representadas como aristas y las intersecciones como nodos, por ello, este archivo contiene una lista de nodos con el *id* y las coordenadas de estos y una lista de aristas con sus identificadores, *ids* de los dos nodos entre los que transitan y datos acerca de la calle, como su velocidad, su longitud, etc. La obtención de este archivo mediante el software JOSM se describió en el apartado 3.1. Generación de la red.
- **Population.xml:** Este archivo contiene la población sintética, es decir, los futuros agentes de la simulación. La población se desarrolla persona a persona, incluyendo la cabecera con los datos básico y la planificación diaria, que se compone de una serie de actividades que se realizan en el día, junto a una serie de desplazamientos realizados para conectar unas actividades con otras, denominados “*leg*”. Este archivo es el que se ha obtenido como salida del sintetizador de población desarrollado en este proyecto.

Una vez que se dispone de estos tres archivos iniciales, se puede iniciar con la simulación. Para ello existen dos opciones, la primera y la recomendada consiste en utilizar un entorno de desarrollo integrado (IDE) como Eclipse para utilizar Matsim. Para ello se clona desde github el proyecto y se importa en Eclipse como un Maven project, esta opción permite al usuario ver el código fuente de Matsim y modificar los *input* desde eclipse. La Figura 41 muestra la interfaz con eclipse con el proyecto de Matsim importado.

Una vez importado el proyecto en Eclipse y guardado el nuevo escenario con los tres archivos XML dentro de la carpeta *scenarios*, simplemente hay que ejecutar el proyecto desde MATSimGUI para que aparezca la interfaz básica que se muestra en Figura42. Donde se elige el archivo config del escenario que se quiera ejecutar para posteriormente ejecutar Matsim.

```

eclipse-workspace - matsim-example-project/src/main/java/org/matsim/gui/MATSimGUI.java - Eclipse IDE
File Edit Source Refactor Navigate Search Project Run Window Help
Package Explorer    MATSimGUI.java    RunMatsim.java    RunMatsimFromExamplesUtils.java    package-info.java
matsim-example-project
src/main/java
org.matsim.gui
  MATSimGUI.java
  package-info.java
org.matsim.project
  RunMatsim.java
  RunMatsimFromExamplesUtils.java
src/test/java
JRE System Library [JavaSE-11]
Maven Dependencies
original-input-data
scenarios
src
target
test
LICENSE
mvnw
mvnw.cmd
pom.xml
README.md
1 * ****
2 * project: org.matsim *
3 *
4 *
5 *
6 * copyright : (C) 2008 by the members listed in the COPYING,
7 *             LICENSE and WARRANTY file.
8 * email    : info at matsim dot org
9 *
10 *
11 *
12 * This program is free software; you can redistribute it and/or modify
13 * it under the terms of the GNU General Public License as published by
14 * the Free Software Foundation; either version 2 of the License, or
15 * (at your option) any later version.
16 * See also COPYING, LICENSE and WARRANTY file
17 *
18 * ****
19
20 package org.matsim.gui;
21
22 import org.matsim.project.RunMatsim;
23
24 public class MATSimGUI {
25
26     public static void main(String[] args) {
27         Gui.show("MATSim GUI from example project", RunMatsim.class );
28     }
29
30     // The jar file is generated by "mvn package". Note that the pom.xml refers at some point to this class here.
31
32 }
33
34

```

Figura 41. Matsim en Eclipse

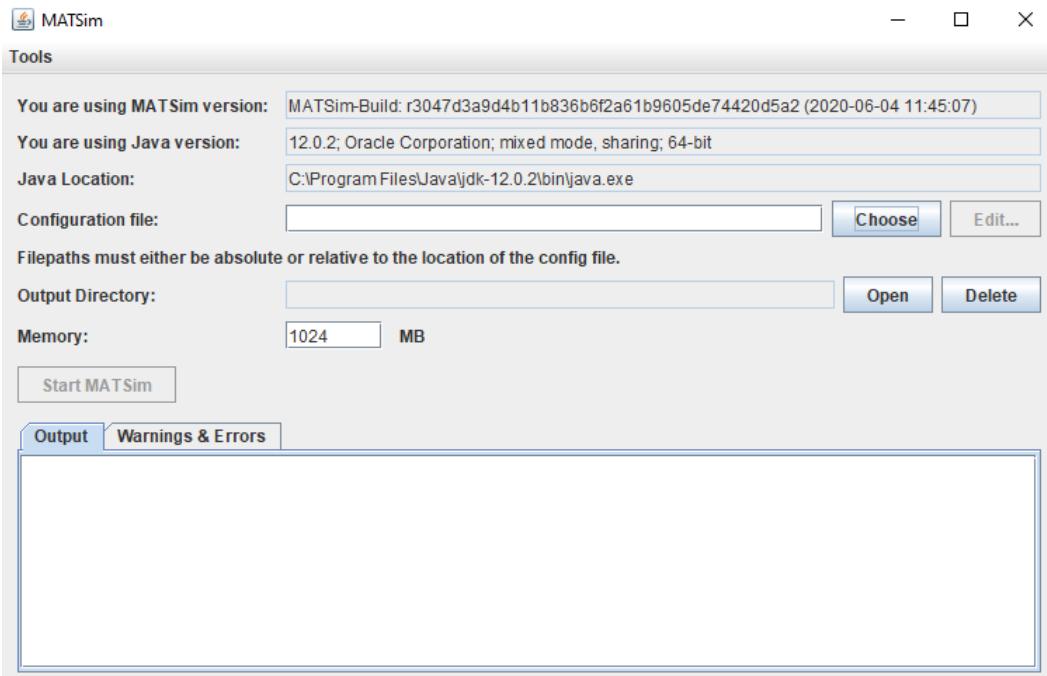


Figura 42. GUI Matsim

La segunda opción para ejecutar Matsim está orientada para los usuarios que no están familiarizados con los ID. Esta versión incluye un archivo jar que, al ejecutarlo, hace aparecer la misma interfaz básica que se obtenía desde eclipse, mostrada en la Figura42, esta opción es la más sencilla, pero admite una menor personalización de la simulación y obliga a editar los archivos *input* para configurar la simulación.

Ejecutada la simulación por cualquiera de los dos métodos, se obtiene una carpeta *output* con los resultados de la simulación. Esta carpeta contiene a su vez otra carpeta con todas las iteraciones simuladas. Cada iteración contiene histogramas con los desplazamientos a lo largo

del día para los distintos modos de transporte. Para cada 10 iteraciones también contiene archivos con los eventos ocurridos y datos de los desplazamientos. Fuera de la carpeta de iteraciones se obtienen los resultados finales, con archivos de las personas, de las actividades, de los eventos, archivos que contienen las puntuaciones para cada una de las iteraciones, etc.

Para más información se recomienda leer el libro de MatSim. Sin embargo, como este es muy extenso e incluye muchos apartados innecesarios para una primera simulación, se recomienda utilizar un extracto de este libro que contiene los capítulos básicos y necesarios para simular, que hace las funciones de guía del usuario [64]

5.1 Simulación del proyecto

Este proyecto se centra en la adquisición de los medios necesarios para la simulación y no tanto en la simulación y experimentación de los desplazamientos en el municipio de Leganés. Aun así, para demostrar la validez de los archivos obtenidos en este proyecto, se va a realizar una simulación inicial con los archivos de población y red viaria de Leganés elaborados a lo largo de este trabajo.

La población utilizada en esta simulación solo contendrá a 500 agentes del fichero original, esto es debido principalmente a tres razones:

- MatSim recomienda ir aumentando la complejidad de las simulaciones progresivamente, para evitar ir acumulando Figues que son difíciles de identificar.
- Via, el programa utilizado para visualizar las simulaciones [65], solo admitía 500 agentes de forma gratuita
- Simular escenarios grandes requiere de muchos recurso, de los que no se disponían.

Para ejecutar MatSim se escogió Eclipse como IDE, obteniendo un fichero output con los resultados de la simulación.

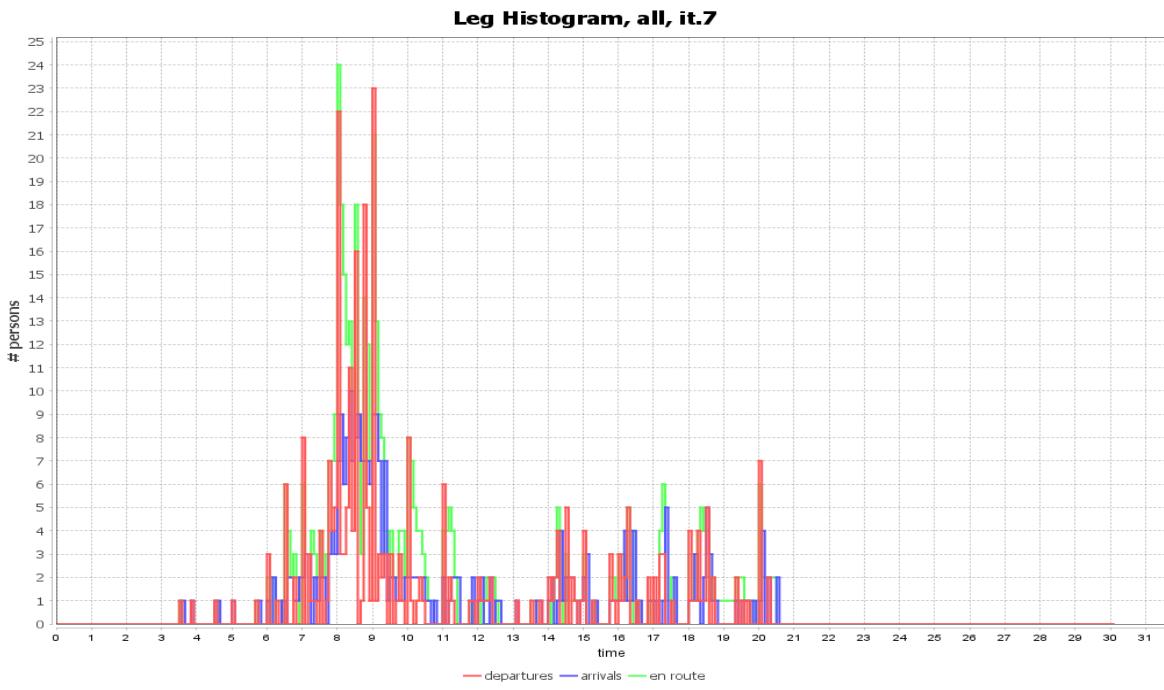


Figura 43. Histograma en la iteración 7

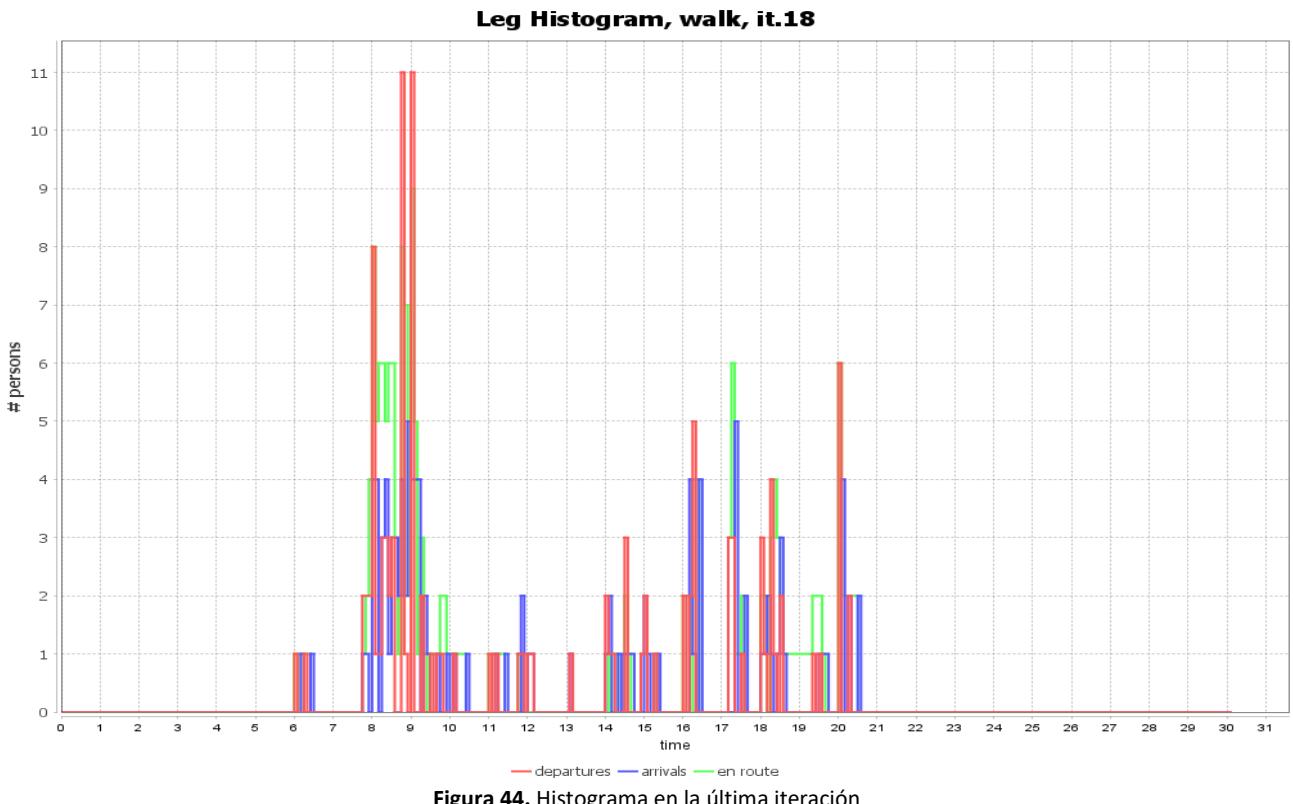


Figura 44. Histograma en la última iteración

En la Figura 43 y la Figura 44 se puede ver la salidas llegadas y transcurso de los desplazamientos a lo largo del día simulado. Por otro lado, en la Figura 45, se pueden ver las puntuaciones mejores, peores y medias por cada iteración, al no configurarse los algoritmos de replanificación los planes no van a ir mejorando sus resultados.

ITERATION	avg. EXECUTED	avg. WORST	avg. AVG	avg. BEST
0	-76.42778271817745	-76.42778271817745	-76.42778271817745	-76.42778271817745
1	-76.42936358570377	-76.42937002469576	-76.42857315194061	-76.42777627918547
2	-76.42781221670253	-76.42937002469576	-76.42853243552877	-76.42777627918547
3	-76.4282630048492	-76.4295470158462	-76.42866763710202	-76.42776152992293
4	-76.42909787131488	-76.4307063660877	-76.42927681074784	-76.42776152992293
5	-76.42808601369875	-76.43103084986352	-76.42943905263574	-76.42776152992293
6	-76.4283072526368	-76.43157657257738	-76.42971191399268	-76.42776152992293
7	-76.42796801959845	-76.43178306225292	-76.42981515883044	-76.42776152992293
8	-76.42880001691515	-76.43237076109845	-76.4301574111761	-76.42768134461826
9	-76.42835419685929	-76.43238551036099	-76.43019334869317	-76.42768134461826
10	-76.42772559240586	-76.43238551036099	-76.43020072332445	-76.42768134461826
11	-76.42908376229374	-76.43370587145324	-76.430831135702	-76.42758641005105
12	-76.42787308503124	-76.43375011924086	-76.43089366412788	-76.42758641005105
13	-76.42787308503125	-76.43375011924086	-76.43094144329125	-76.42758641005105
14	-76.42799107913152	-76.43375011924086	-76.43100642992758	-76.42758641005105
15	-76.4289466460056	-76.43459413100662	-76.43147514180848	-76.42760115931358
16	-76.42845100270506	-76.43459413100662	-76.43154227477547	-76.42758641005105
17	-76.42791089382686	-76.43459413100662	-76.43156464449034	-76.42755691152597
18	-76.42980106931128	-76.43497761183258	-76.43205420667775	-76.42755691152597
19	-76.42815976001143	-76.43497761183258	-76.43213059096318	-76.42755691152597
20	-76.42882014684663	-76.43601960821519	-76.43267002573266	-76.42755691152597

Figura 45. Puntuaciones por iteración

Para terminar, en la Figura 46 se puede apreciar gráficamente la simulación de 24 hora de Leganés, los puntos representan a las personas que se desplazan andando y las flechas representan a los agentes que se desplazan en coche a sus respectivos destinos para realizar una actividad.

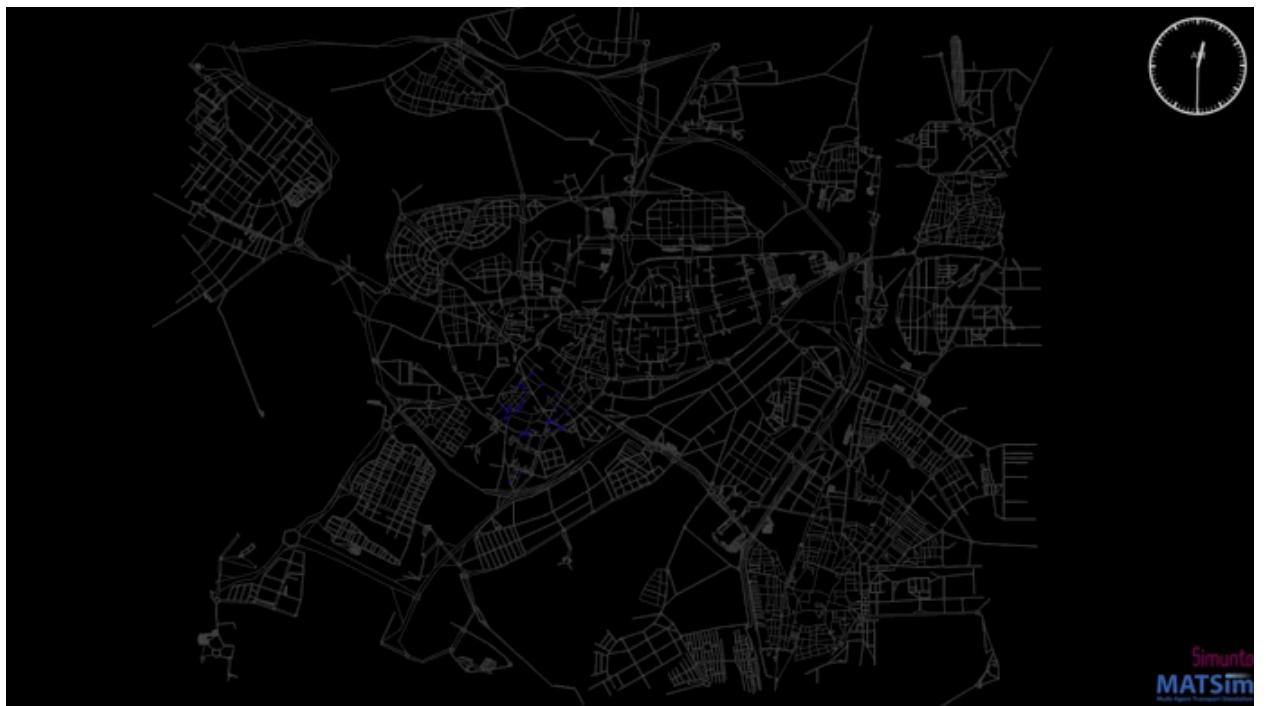


Figura 46. Simulación de Leganés en Matsim

El video con la simulación completa se puede ver en YouTube en el siguiente [enlace](#)

6. Gestión del proyecto

En este apartado se presenta de la planificación del proyecto, se desglosará el presupuesto necesario para el desarrollo del proyecto y, por último, se introducirá el impacto socio-económico esperado tras la elaboración de este trabajo.

6.1 Planificación

Para arrojar luz sobre los tiempos dedicadas a cada una de las tareas realizadas, se ha elaborado una tabla donde se puede visualizar la progresión del proyecto. El análisis de los datos expuestos en la misma permite comprobar como algunas de las tareas se solapan en el tiempo, otras tareas como la programación del Familiador se ha extendido a lo largo del proyecto por la inclusión de pequeñas mejoras hacia el final del trabajo.

Tareas	Fecha de inicio	Fecha de Fin	Duración en Horas
Estudio del proyecto	13/01/21	13/09/21	169
Estudio de Matsim	13/01/21	20/01/21	28
Estudio sobre la generación de la red viaria	22/01/21	24/01/21	17
Estudio sobre la generación de la población	28/01/21	16/02/21	59
Estudio sobre la generación de planes diarios	02/02/21	13/09/21	36
Estudio de sintetizadores de población	05/02/21	22/02/21	29
Análisis del problema	22/02/21	26/02/21	23
Diseño de la solución	5/03/21	22/12/21	338
Búsqueda de fuentes de datos	28/02/21	11/03/21	23
Preprocesado de datos censo	24/03/21	24/03/21	3
Generación del mapa	13/03/21	16/03/21	22
Preprocesado datos INE	28/03/21	24/07/21	54
Preprocesamiento datos Catastro	05/03/21	16/11/21	14
Programación Familiador	12/03/21	13/12/21	126
Programación Planificador	09/09/21	20/12/21	91
Análisis del sistema	19/10	22/12/21	5
Redacción de la memoria	15/07/21	27/02/22	378
Introducción	28/01/22	10/02/22	7

Estado del arte ciudades y transportes	22/07/21	30/01/22	45
Estado del arte Teoría agentes	15/07/21	24/10/21	56
Estado del arte MatSim	13/10/21	15/10/21	16
Obtención de los datos	20/07/21	14/01/22	23
Familiarizador	26/10/21	19/01/22	77
Planificador	18/12/21	27/01/22	79
Simulación/Conclusiones	04/02/22	13/02/22	21
Revisiones	28/12/21	27/02/22	54

Tabla 35. Planificación

Para tener una mejor perspectiva del proyecto y que se pueda ver visualmente la duración de cada apartado, se incluye un diagrama de Gantt dividido por meses en la Figura 47.

El proyecto se ha desarrollado a lo largo de un año y dos meses, la implementación del programa ha durado en torno a un tercio del trabajo, aunque el porcentaje de tiempo invertido en esta parte se esperaba superior, el trabajo de investigación realizado para poder llegar a programar ha sido muy superior al tiempo estimado, tanto de los programas, como de la búsqueda y del preprocesado de datos.

6.2 Presupuesto

En este apartado se hace un análisis de los recursos y costes de materiales, personales y de herramientas para el desarrollo del proyecto.

6.2.1 Costes del personal

Para la elaboración de cualquier proyecto siempre es necesario formar un equipo de personas con diferentes perfiles y conocimientos. Uno de los profesionales debería asumir la función de jefe de proyecto y de coordinación de los demás miembros, su función consistiría en planificar, gestionar y supervisar. Para el desarrollo del software será necesario contratar un ingeniero informático junior que tenga conocimientos de Python y de bases de datos para programar el software y para poder realizar las consultas a las bases de datos. Finalmente dada la naturaleza del proyecto se requerirá la participación de un sociólogo que aporte conocimientos en torno a la composición y el desplazamiento de los hogares.

Los salarios anuales de los profesionales han sido obtenidos del buscador de empleo Indeed, que en base a sueldos reales anónimos obtiene una media de los salarios en España [66]–[68] y los costes por hora se han obtenido siguiendo la siguiente formula:

$$\text{Coste por hora} = \text{Coste total} / (140 \text{ horas/mes} * 14 \text{ meses})$$

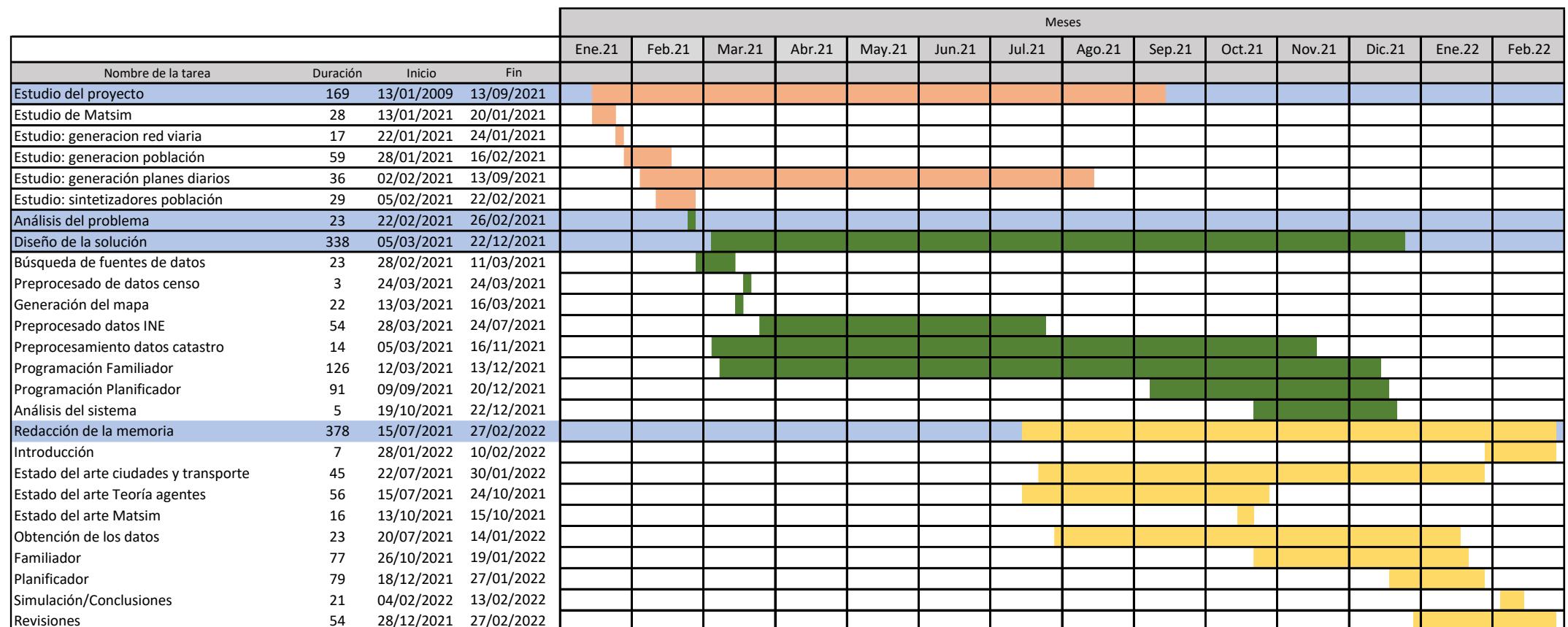


Figura 47. Diagrama de Gantt

Profesional	Salario anual	S.S. anual 29.8%[69]	Coste total	Horas	Coste por hora	Coste atribuible
Jefe del proyecto	35.709€	9.998,52€	45.707,52€	138	23,32€	3.218,16€
Ingeniero informático	25.807€	7.690,49€	33.497,49€	487	17,09€	8.322,83€
Sociólogo	17.520€	5.220,96€	22.740,96€	283	11,60€	3.282.80€
Total gastos de personal		14.823,79€				

Tabla 36. Costes de personal

6.2.2 Costes del equipo

En este apartado se contabilizan los gastos del proyecto en hardware y software necesarios para la realización de este.

Producto	Coste total	Uso en meses	Vida útil en meses	Coste para el proyecto
Hardware				
Ordenador x2	1200€	14	48	350€ 700€ (x2)
Monitor 23' x2	139€	14	60	32,43€ 64,86€ (x2)
Software				
JOSM	Gratis	14	-	-
Python	Gratis	14	-	-
Pycharm	Gratis	14	-	-
R	Gratis	14	-	-
RStudio	Gratis	14	-	-
Matsim	Gratis	14	-	-
Total del equipo		764,86€		

Tabla 37. Costes del equipo

6.2.3 Costes para fungibles y otros gastos varios

Los gastos del material fungible y otros gastos comunes se recogen en este apartado.

Producto	Coste
Luz y agua	50€ 700€ (x14)
Conexión a internet	35€ 490€ (x14)
Material de oficina	30€
Total de fungibles y otros	1220€

Tabla 38. Costes de fungibles y otros

6.2.4 Costes totales del proyecto

Este último apartado contiene los costes finales teniendo en cuenta todos los gastos anteriormente mencionados para la elaboración de este proyecto.

Producto	Coste
Coste del personal	14.823,79€
Coste del equipo	764,86€
Coste de fungibles y otros	1220€
IVA 21%	2203€
Total del proyecto	19.011,65€

Tabla 39. Costes totales del proyecto

6.3 Impacto socio-económico

Como se ha podido ver con la extensa argumentación del estado del arte, la sociedad está viviendo cambios drásticos, la población mundial sigue creciendo, al menos hasta el año 2060. Se están alcanzando unas cifras insostenibles que el planeta no puede permitirse. La migración tiende a la desruralización, en el campo la falta de servicios, los menores ingresos y la falta de oportunidades generan un flujo de gente que decide abandonar el campo para labrarse un nuevo futuro en las urbes.

La suma del aumento de la población mundial unido a la llegada de población rural a las ciudades está produciendo una masificación de estas. Esta ciudades, cada vez más congestionadas, acumulan problemas y muestran un paisaje en el que, entre múltiples aspectos a replantear, los transportes, la población y sus necesidades de desplazamiento precisan, especialmente, de una reflexión y una intervención urgente.

La contaminación y el calentamiento global, la saturación de vehículos de transporte en las ciudades, la existencia de multitud de espacios destinados a contener y permitir el desplazamiento de estos vehículos en las urbes y el uso de millones de horas de los ciudadanos dedicadas a gestionar su transporte de unos lugares a otros, la cantidad de accidentes de tráfico, y un largo etcétera de fenómenos como los enumerados, que proporcionan una buena descripción de la realidad actual.

El mundo, y especialmente Europa, está pasando de la toma de conciencia de esta realidad a acuñar soluciones que permitan ir afrontando la mejora de los aspectos más preocupantes. Es aquí donde este trabajo puede tener un impacto social y económico importante, pues todas estas nuevas ideas y proyectos necesitan ser contrastados, necesitan saber si son útiles para la mejora de las condiciones iniciales, si son viables económicamente, si tienen un buen impacto ecológico. Para ello es imprescindible la simulación, dado que la implementación directa de estos proyectos es inviable por los altos costos. Una simulación gracias a la cual se va a arrojar luz sobre todas estas cuestiones, para una vez resultas, poder llegar a la implementación. Sin embargo, acceder a la simulación es una tarea compleja, como se ha constatado al inicio de este trabajo por la cantidad de horas y de esfuerzo que ha supuesto. La escasez de manuales y recursos fáciles e intuitivos para simular, crean una barrera para la iniciación y desarrollo de proyectos de esta índole, dificultando o impidiendo el acceso de futuros profesionales al desarrollo de simulaciones.

Para paliar esta situación y eliminar las barreras, este trabajo prepara y configura un entorno de simulación de transporte urbano basado en agentes, facilitando con esto el trabajo a futuros usuarios. Este proyecto aplica un nuevo enfoque para la generación de poblaciones sintéticas basado en parámetros sociológicos, que permitan crear poblaciones realistas con la que simular.

Una simulación es un proceso aplicable a múltiples campos de la economía y la mejora de las condiciones sociales de la población que permitirá, con un gasto moderado, poder realizar un estudio virtual que, en si se pretendiera llevar a la práctica tendría unos costes extraordinarios o moriría antes de haber nacido desechándose.

Pensamos que una contribución a la mejora y aumento del uso de la simulación va a tener unas amplias consecuencias mejorando la probabilidad de que multitud de proyectos que mejorarían las expectativas socioeconómicas en el marco sobre el que se apliquen, salgan adelante con las ventajas que ello supondrá.

En el caso de que estas simulaciones tuvieran como resultado la puesta en práctica de servicios de vehículos autónomos, las consecuencias socioeconómicas en todos los sectores relacionados con la automoción, el transporte y el desplazamiento de ciudadanos de un lugar a otro serían inmensas. Se darían cambios en el tipo y la cantidad de vehículos producidos, la transformación de miles de empleos o la planificación del tráfico y los desplazamientos. Los tiempos de desplazamiento y la contaminación generada por los vehículos disminuiría, junto con el número de accidentes de tráfico y sus consecuencias humanas y económicas.

Como se puede apreciar, la simulación como forma de análisis de la viabilidad socioeconómica de múltiples proyectos, debería tener una alta incidencia en el mercado laboral, empresarial y social.

7. Conclusiones y trabajos futuros

Este apartado concluye este trabajo con un análisis del proyecto realizado, para dar a conocer todas las conclusiones extraídas del mismo, justificando el cumplimiento de los objetivos propuestos. En un segundo punto se van a reflejar las líneas de trabajo futuras para mejorar y ampliar el programa aquí desarrollado.

7.1 Conclusiones

El objetivo principal de este proyecto era el de la preparación y configuración de un entorno de simulación, exponiendo y desarrollando todos los procesos necesarios previos a la simulación.

Durante los distintos apartados de este trabajo se han ido desarrollando los conocimientos básicos necesarios, definiendo los principios básicos de los agentes para luego exponer los sistemas multiagente y la simulación basada en agentes. Se ha introducido Matsim, el simulador utilizado en este proyecto, del que se han detallado su funcionamiento básico y sus componentes, exponiendo las razones que han llevado a la creación de una población y de una red viaria para poder simular.

Posteriormente, se desarrollaron las distintas formas existentes para la obtención de los archivos necesarios para simular. Se han especificado los datos esenciales y la forma de adquirirlos en este proyecto, nombrando otras formas mediante las cuales también se pueden obtener los archivos de población y de la red viaria. Una vez expuesto el funcionamiento del sintetizador de población desarrollado, se describió la forma en que ejecutar Matsim, mostrando que los ficheros obtenidos como salida de *The Synthetic Population* funcionaban correctamente. Una vez llegados a este punto se puede afirmar que el objetivo principal del trabajo quedó cumplido.

El segundo objetivo del proyecto consistía en desarrollar un nuevo método para la generación de poblaciones sintéticas basado en la sociología, un enfoque distinto, que se centra en estudiar los factores que definen la composición y los desplazamientos de los hogares.

Al término de este trabajo se ha conseguido desarrollar un software capaz de generar una población sintética completa de una región deseada, a partir de datos estadísticos de la población objetivo. El archivo obtenido como salida de este software tiene el formato correcto y funciona en Matsim.

Los resultados del programa fueron valorados en el apartado 4.3 Evaluación, donde se compararon las familias creadas con las reales y el total de los tipos de viajes asignados con los viajes reales de la encuesta de movilidad. Los porcentajes en la mayoría de las categorías eran muy similares, aunque con algunas excepciones. Dado el estado inicial de desarrollo del programa, se considera que los resultados obtenidos son satisfactorios ya que se va a seguir implementando nuevas tipología en prototipos posteriores con los que se espera que los resultados mejoren.

La principal dificultad de este trabajo ha radicado en la complejidad, la inmediatez y la no-linealidad típica de muchos fenómenos sociales. Es un reto lograr representar relaciones sociales que tienen asociada la aleatoriedad y la irracionalidad, pues esta hace que no siempre se siga la solución óptima. Esta complejidad humana hace que, para conseguir unos resultados que se ajusten a la realidad, sea necesario tener en cuenta una gran cantidad de variables, que no siempre están disponibles. Cada vez que se entra en una nueva fase de proyecto, se comprueba

que las ideas simples que se tenían de origen distan mucho de la realidad, a medida que se profundiza en cada una de las ideas, se van abriendo nuevos campos de investigación. Conseguir una representación fiable no es una tarea trivial, es necesario estudiar muchos campos distintos para poder trabajar, se entiende por tanto, que este proyecto tiene muchos caminos por los que se puede seguir desarrollando y de los que se espera que sean recorridos en un futuro próximo.

7.2 Trabajos futuros

Se espera que este proyecto abra un camino para la creación de poblaciones sintéticas a través de la sociología. El software desarrollado planteado una posible solución a través de este nuevo enfoque, no obstante, existen multitud de caminos distintos para abordar la generación de personas a partir de las estructuras sociales, lo que deja una puerta abierta a otros investigadores que deseen continuar el software desarrollado o abordar la creación de nuevos programas que tengan como base las ideas planteadas en este Trabajo Fin de Grado.

El programa desarrollado en este trabajo es un primer prototipo que, debido a su complejidad, no contiene todas las ideas planteadas a lo largo de este documento. Todas estas ideas de mejora del programa se han ido desarrollando a lo largo de la descripción de este y especialmente fueron propuestas en los apartados de mejoras, tanto del Familiador, como del Planificador, que aparecen en los apartados 3.2.3.5 y 3.2.4.7 del trabajo respectivamente. Se espera que la tarea aquí recogida facilite la labor de otros investigadores que deseen continuar el proyecto, pues se han dejado indicados todos los puntos posibles de mejora.

Como anhelo, se busca que el trabajo aquí desarrollado tenga alguna repercusión futura en el campo de los sintetizadores de población y que las puertas abiertas por este proyecto consigan abrir otras puertas para facilitar la simulación y poder abordar los retos que propone el crecimiento actual de las ciudades.

Bibliografía

- [1] "GNU General Public License v2.0 - GNU Project - Free Software Foundation," *GNU General Public License, version 2*, 1991. <https://www.gnu.org/licenses/old-licenses/gpl-2.0.html> (accessed Feb. 03, 2022).
- [2] "Database Contents License (DbCL) v1.0 — Open Data Commons: legal tools for open data," *Open Data Commons*. <https://opendatacommons.org/licenses/dbcl/1-0/> (accessed Feb. 03, 2022).
- [3] D. Herranz, "El riesgo de colapso demográfico se concentra en Europa," *Público*, 2020. <https://www.publico.es/economia/riesgo-colapso-demografico-concentra-europa.html> (accessed Aug. 27, 2021).
- [4] R. BBC, "El dramático descenso de los nacimientos en el mundo (y cuáles serán sus profundas consecuencias)," *BBC News Mundo*, 2020. <https://www.bbc.com/mundo/noticias-53417504> (accessed Aug. 27, 2021).
- [5] OCU, "Un estudio señala que el coste del primer año de vida de un hijo asciende 7.706 euros," *OCU*, 2020. <https://www.ocu.org/organizacion/prensa/notas-de-prensa/2020/coste-hijo240920> (accessed Aug. 30, 2021).
- [6] I. Alloza, "¿Cuál será la población mundial en 2050?," *Geografía Infinita*. <https://www.geografiainfinita.com/2020/11/cual-sera-la-poblacion-mundial-en-2050/> (accessed Aug. 29, 2021).
- [7] J. A. Ortega, "Horizonte demográfico de Asia," no. 1, 2004.
- [8] R. Gómez and J. Moya, "África y el suicidio demográfico," *Atalayar*, 2019. <https://atalayar.com/content/áfrica-y-el-suicidio-demográfico> (accessed Sep. 01, 2021).
- [9] J. Bongaarts and J. Casterline, "Fertility Transition: Is sub-Saharan Africa Different?," *Popul. Dev. Rev.*, vol. 38, no. SUPPL.1, pp. 153–168, 2013, doi: 10.1111/j.1728-4457.2013.00557.x.
- [10] S. Emil Vollset *et al.*, "Fertility, mortality, migration, and population scenarios for 195 countries and territories from 2017 to 2100: a forecasting analysis for the Global Burden of Disease Study," *Glob. Heal. Metrics* www.thelancet.com, 2020, doi: 10.1016/S0140-6736(20)30677-2.
- [11] Demographia, "Demographia World Urban Areas (Built-Up Urban Areas or World Agglomerations) 10 th Annual Edition May 2014 Revision," no. March, p. 129, 2014, [Online]. Available: <http://demographia.com/db-worldua.pdf>.
- [12] UN, "Las ciudades seguirán creciendo, sobre todo en los países en desarrollo," *UN*, 2018. <https://www.un.org/development/desa/es/news/population/2018-world-urbanization-prospects.html> (accessed Sep. 02, 2021).
- [13] Y. Ariani, P. Soeharso, and D. R. Sjarif, "Genetics and genomic medicine in Indonesia," doi: 10.1002/mgg3.284.
- [14] S. Dustdar, S. Nastic, and O. Šcekić, "Smart cities: The internet of things, people and systems," *Smart Cities Internet Things, People Syst.*, pp. 1–268, May 2017, doi: 10.1007/978-3-319-60030-7.
- [15] M. E. Z. Rodríguez, G. Grado, D. E. E. Tutora, C. Fran, O. Marie, and G. Fecha, "SMART CITIES : ¿ La Smart City solventa los problemas futuros que se estiman ? ¿ Es la única opción ?, " 2021.
- [16] Leaving Babylon, "Being there," *Leaving Babylon*. <https://leavingbabylon.wordpress.com/book/being-there/> (accessed Sep. 04, 2021).
- [17] VIU, "Smart mobility: concepto y principios," *VIU*, 2019. <https://www.universidadviu.com/es/actualidad/nuestros-expertos/smart-mobility-concepto-y-principios> (accessed Sep. 04, 2021).
- [18] Smartrak, "Why Mobility as a Service is the future of transportation," *Smartrak*. <https://smartrak.com/maas-the-future-of-transportation/> (accessed Feb. 25, 2022).

- [19] Juan Ramirez, "Historia de los vehículos autónomos - Los vehículos autónomos." <https://sites.google.com/site/losvehiculosautonomos/historia-de-los-vehiculos-autonomos> (accessed Jul. 30, 2021).
- [20] J. A. Valero-Matas and A. De La Barrera, "The Autonomous Car: A better future?*", *Sociol. y Tecnociencia*, vol. 10, no. 1, pp. 136–159, 2020, [Online]. Available: https://search.proquest.com/docview/2356820386/accountid=7098%0Ahttps://ttu-primo.hosted.exlibrisgroup.com/openurl/01TTU/01TTU?genre=article&issn=&title=The+Autonomous+Car%3A+A+better+future%3F*&volume=10&issue=1&date=2020&atitle=The+Autonomous+Car%3A+A+b.
- [21] M. HERRÁEZ, "Honda lleva el coche autónomo al siguiente nivel," *Motor EL PAÍS*, 2021. <https://motor.elpais.com/tecnologia/honda-coche-autonomo-siguiente-nivel/> (accessed Feb. 07, 2022).
- [22] Nikkei, "Japan to create legal framework for level 4 self-driving cars," *Nikkei Asia*, 2021. <https://asia.nikkei.com/Business/Automobiles/Japan-to-create-legal-framework-for-level-4-self-driving-cars> (accessed Jan. 05, 2022).
- [23] V. Fuentes, "Reino Unido dará luz verde al coche autónomo de nivel 3 este mismo año: comienza su regulación," *Motorpasion*, 2021. <https://www.motorpasion.com/seguridad/reino-unido-dara-luz-verde-al-coche-autonomo-nivel-3-este-ano-comienza-su-regulacion> (accessed Jan. 05, 2022).
- [24] C. Otero, "Los robotaxis ya circulan por China: Taxis autónomos sin conductor," *AS*, 2021. https://as.com/meristation/2021/03/16/betech/1615902980_825205.html (accessed Feb. 25, 2022).
- [25] J. M. Corchado and J. M. M. López, "Introducción a la teoría de agentes y sistemas multiagente," pp. 1–17, 2002.
- [26] HISOUR Arte Cultura Historia, "Arquitectura de la subsunción," *HISOUR Arte Cultura Historia*. <https://www.hisour.com/es/subsumption-architecture-43228/> (accessed Feb. 02, 2022).
- [27] "Multi-agent system," *Wikipedia*. https://en.wikipedia.org/wiki/Multi-agent_system (accessed Feb. 02, 2022).
- [28] V. Juan, B. Navarro, and A. G. Boggino, "Un sistema Multi-agente para mercados de logística inversa," Universitat Politècnica de València, 2018.
- [29] J. A. J. BUILES, D. A. O. CARRANZA, and J. F. O. GÓMEZ, "SMART: SISTEMAS MULTI-AGENTE ROBÓTICO," *DYNA*, vol. 75, no. 154, pp. 179–186, Jan. 2008, Accessed: Feb. 02, 2022. [Online]. Available: <https://revistas.unal.edu.co/index.php/dyna/article/view/1726>.
- [30] V. Dávila, "Simulación multi-agente para gestión de desastres y reducción de riesgo Multi-agent simulation for disaster management and risk reduction," pp. 21–30, 2011, [Online]. Available: <http://erevistas.saber.ula.ve/index.php/cienciaeingenieria/article/viewArticle/3868>.
- [31] A. R. Abed, B. Cruz, and L. J. E. Nieto, "Sistema multiagente para la simulación de la defensa de un edificio ante un ataque enemigo," 2006.
- [32] V. J. B. Navarro and A. G. Boggino, "Aplicaciones Industriales de los Sistemas Multiagente," *Univ. Politécnica Val.*, Accessed: Feb. 02, 2022. [Online]. Available: <http://imsg.enme.ucalgary.ca/publication/abm.htm>:
- [33] M. Niazi and A. Hussain, "Agent-based computing from multi-agent systems to agent-based models: A visual survey," *Scientometrics*, vol. 89, no. 2, pp. 479–499, 2011, doi: 10.1007/s11192-011-0468-9.
- [34] J. I. G. V. Medina, "Agent-based modelling: A new way of exploring social phenomena," *Rev. Esp. Investig. Sociol.*, no. 136, pp. 91–110, 2011, doi: 10.5477/cis/reis.136.91.
- [35] "Agent-based modeling – Institute for Transport Planning and Systems | ETH Zurich," *ETH Zurich*. <https://www.ivt.ethz.ch/en/research/matsim.html> (accessed Feb. 03, 2022).

- [36] A. Horni, K. Nagel, and K. W. Axhausen, *The Multi-Agent Transport Simulation Title of Book : The Multi-Agent Transport Simulation MATSim*. 2016.
- [37] A. Neumann and M. Zilske, *MATSim JOSMNetwork Editor*. 2018.
- [38] M. Raifer, “overpass turbo,” *overpass turbo*. <https://overpass-turbo.eu/> (accessed Feb. 03, 2022).
- [39] W. Schneider, “BBBike extracts OpenStreetMap,” 2021. <https://extract.bbbike.org/> (accessed Feb. 08, 2022).
- [40] “Planet OSM.” <https://planet.openstreetmap.org/> (accessed Feb. 08, 2022).
- [41] S. Coast, “OpenStreetMap,” 2004. <https://www.openstreetmap.org/#map=17/40.33084/-3.76361> (accessed Feb. 03, 2022).
- [42] “Josm Matsim plugin.” <https://github.com/matsim-org/josm-matsim-plugin> (accessed Mar. 10, 2021).
- [43] N. Kühnel, “Network Editor for JOSM.”
- [44] C. Zhuge, C. Shao, S. Wang, and Y. Hu, “An agent-and gis-based virtual city creator: A case study of Beijing, China,” *J. Transp. Land Use*, vol. 11, no. 1, pp. 1231–1256, 2018, doi: 10.5198/jtlu.2018.1270.
- [45] R. J. Beckman, K. A. Baggerly, and M. D. McKay, “Creating synthetic baseline populations,” *Transp. Res. Part A Policy Pract.*, vol. 30, no. 6 PART A, pp. 415–429, 1996, doi: 10.1016/0965-8564(96)00004-3.
- [46] B. Farooq, M. Bierlaire, R. Hurtubia, and G. Flötteröd, “Simulation based population synthesis,” *Transp. Res. Part B Methodol.*, vol. 58, pp. 243–263, 2013, doi: 10.1016/j.trb.2013.09.012.
- [47] S. Felbermair, F. Lammer, E. Trausinger-Binder, and C. Hebenstreit, “Generating synthetic population with activity chains as agent-based model input using statistical raster census data,” *Procedia Comput. Sci.*, vol. 170, pp. 273–280, 2020, doi: 10.1016/j.procs.2020.03.040.
- [48] G. D. Floriana Gargiulo, Sonia Ternes, Sylvie Huet, “An Iterative Approach for Generating StatisticallyRealistic Populations of Households.” 2010.
- [49] H. Bar-Gera, K. C. Konduri, B. Sana, X. Ye, and R. M. Pendyala, “Estimating Survey Weights with Multiple Constraints Using Entropy Optimization Methods,” *Transp. Res. Board 88th Annu. Meet.*, vol. 6850, no. November, p. 22, 2009.
- [50] A. Societies and S. Simulation, “Kirk Harland , Alison Heppenstall , Dianna Smith and Mark Birkin (2012) Creating Realistic Synthetic Populations at Varying Spatial Scales : A Comparative Critique of Population Synthesis Techniques Spatial microsimulation algorithms,” *Simulation*, vol. 15, no. 2012, pp. 1–24, 2012.
- [51] L. Ma and S. Srinivasan, “Synthetic population generation with multilevel controls: A fitness-based synthesis approach and validations,” *Computer-Aided Civil and Infrastructure Engineering*, vol. 30, no. 2. pp. 135–150, 2015, doi: 10.1111/mice.12085.
- [52] A. S. Chandra R. Bhat, Jessica Y. Guo, Sivaramakrishnan Srinivasan, “CEMDAP,” 2004. <https://www.caee.utexas.edu/prof/bhat/CEMDAP.htm> (accessed Jan. 11, 2022).
- [53] B. M. Paul, J. Doyle, B. Stabler, J. Freedman, and A. Bettinardi, “Population Synthesizer Development – PopulationSim - RSG,” 2018. <https://rsginc.com/project/population-synthesizer-development/> (accessed Jan. 11, 2022).
- [54] MARG, “PopGen: Synthetic Population Generator [online]. Mobility Analytics Research Group.,” MARG, 2016. <https://www.mobilityanalytics.org/popgen.html> (accessed Feb. 26, 2022).
- [55] Deloitte and IPD, “Encuesta domiciliaria de movilidad de la Comunidad de Madrid 2018,” p. 76, 2020, [Online]. Available: <https://www.crtm.es/conocenos/planificacion-estudios-y-proyectos/encuesta-domiciliaria/edm2018.aspx>.
- [56] A. Delgado, “Cálculo de rutas optimas para coches eléctricos en entornos urbanos.”
- [57] INE, “Encuesta Continua de Hogares,” INE.

- <https://www.ine.es/dynt3/inebase/index.htm?type=pcaxis&path=/t20/p274/serie/prov/p02&file=pcaxis&L=0&dh=0&capsel=0> (accessed Feb. 03, 2022).
- [58] P. Miret Gamundi, “La similitud entre los componentes de las parejas jóvenes en España en la primera década del siglo XXI ¿Cada vez más iguales?,” *Rev. Estud. Juv.*, vol. 90, pp. 225–255, 2009.
- [59] M. R. C. D. S. J. M. S. M. L. I. M. Pérez, “Plan de movilidad urbana sostenible de Leganés,” *Director*, p. 3367264, 2010.
- [60] INE, “Estudio de movilidad de la población a partir de datos de telefonía móvil (EM-4) julio y agosto de 2021,” vol. 4, pp. 1–7, 2021.
- [61] J. M. Agudo Benavides, M. A. Martinez Cardoso, and W. Sierra-Barón, “Caracterización de la población que asiste a casinos en la ciudad de Neiva,” *Rev. Estud. Psicosoc. Latinoam.*, vol. 1, pp. 8–15, 2018, doi: 10.25054/26196077.1614.
- [62] D. Ziemke, I. Kaddoura, and K. Nagel, “The MATSim open Berlin scenario: A multimodal agent-based transport simulation scenario based on synthetic demand modeling and open data,” *Procedia Comput. Sci.*, vol. 151, no. 2018, pp. 870–877, 2019, doi: 10.1016/j.procs.2019.04.120.
- [63] F. Poletti, “Public Transit Mapping on Multi-Modal Networks in MATSim,” *Strasse und Verkehr*, no. 7–8, pp. 22–25, 2017, [Online]. Available: <https://ethz.ch/content/dam/ethz/special-interest/baug/ivt/ivt-dam/publications/students/501-600/sa530.pdf> <https://www.ethz.ch/content/dam/ethz/special-interest/baug/ivt/ivt-dam/news/vss-poletti-matsim.pdf>.
- [64] K. W. Axhausen, “The Multi-Agent Transport Simulation MATSim (Extract of selected chapters to serve as user guide),” *Multi-Agent Transp. Simul. MATSim*, p. 88, 2016, doi: 10.5334/baw.
- [65] “Simunto Via,” *Simunto*. <https://simunto.com/via/> (accessed Feb. 14, 2022).
- [66] Indeed, “Salario de un Jefe de proyecto en España,” *Indeed*. <https://es.indeed.com/career/jefe-de-proyecto/salaries> (accessed Feb. 04, 2022).
- [67] Indeed, “Salario de un Ingeniero informático en España,” *Indeed*. <https://es.indeed.com/career/ingeniero-informatico/salaries> (accessed Feb. 04, 2022).
- [68] Indeed, “Salario de un Sociólogo en España,” *Indeed*. <https://es.indeed.com/career/sociologo/salaries> (accessed Feb. 04, 2022).
- [69] El Economista, “Cuánto cotiza un trabajador a la Seguridad Social: esto es lo que debe aparecer en su nómina,” *El Economista*, 2021. <https://www.eleconomista.es/economia/noticias/11433029/10/21/Cuanto-cotiza-un-trabajador-a-la-Seguridad-Social-esto-es-lo-que-debe-aparecer-en-su-nomina.html> (accessed Feb. 04, 2022).
- [70] INE, “Censos de Población y Viviendas 2011,” 2011. <https://www.ine.es/censos2011/tabcenso/Inicio.do> (accessed Mar. 20, 2022).
- [71] Catastro, “Fichero informático de remisión de catastro (bienes inmuebles urbanos, rústicos),” 2021.

Anexo A

A.1. Datos y Fuentes

En este apartado se van a enumerar todas las fuentes de datos utilizadas, exponiendo como se obtuvieron los datos de estas fuentes y desarrollando el preprocesado que se tuvo que realizar a los datos para que The Synthetic Population pudiera utilizarlos.

A.1.1. Censo de Leganés

Origen y selección de datos

El censo de Leganés fue descargado de la página del ayuntamiento¹², donde se podían encontrar los datos del censo en dos formatos, por distritos o por barrios. Inicialmente se decide tomar los datos por barrios, por su mayor precisión al trabajar con subdivisiones menores, sin embargo, estos presentaban los siguientes problemas:

- Los datos por barrio no estaban completos. Por ejemplo, el barrio de Campo Claro, que aparecía en el mapa de estadísticas de la página del ayuntamiento, no tenía su archivo asociado en el censo.
- Los datos ofrecidos por el INE a nivel de barrio eran datos obsoletos del 2011, pues el censo de población y viviendas a estos niveles se realiza cada 10 años y el de 2021 no será publicado hasta 2022.
- De haber realizado un estudio barrio a barrio, se hubiera añadido una mayor complejidad al proyecto.
- Los microdatos ofrecidos por el catastro en el caso de Leganés solo estaban clasificados hasta el nivel de distrito, y aunque existía un campo con subdivisiones menores de los municipios, este se encontraba vacío. Por otro lado, se pensó en utilizar la columna de los códigos postales, sin embargo, se comprobó que las áreas de los códigos postales no coincidían con las áreas de los barrios del censo de Leganés.

Los problema de los datos por barrios enumerados limitaban la capacidad del proyecto para abordar la siguiente parte de planificación. Como el objetivo de este trabajo, era el de guiar a los lectores hacia la simulación con Matsim, se decidió utilizar los datos censales por distrito y así no limitar el alcance del proyecto.

Preprocesado de los datos

El preprocesado del censo variará en función del formato en que se encuentren los datos. En el caso particular del ayuntamiento de Leganés, los datos del censo descargados venían en forma de tabla en pdf. El archivo se transformó a un csv dejando solo los datos relativos al número de personas para cada edad, divididas por sexo y ordenadas de menor a mayor edad, quedando como resultados 7 archivos csv, uno por cada distrito de Leganés. Una vez preprocesados los datos, estos ya estaban listos para que el sintetizador los pudiera leer y utilizar.

¹² Página Web del ayuntamiento de [Leganés](#)

A.1.2. Datos estadísticos poblacionales del INE

Fuentes iniciales de datos

Inicialmente se buscaron múltiples fuentes de datos, como por ejemplo, el portal de datos abiertos del Ayuntamiento de Madrid, que tiene disponibles un gran número de conjuntos de datos de temas muy variados. También se encontraron distintas páginas web que contenían resúmenes estadísticos sacados a partir de datos de terceros, tras analizar estas páginas se determina que los datos carecían del rigor suficiente para ser incluidos en este trabajo.

Tras comprobar la cantidad y calidad de los datos obtenidos de las distintas fuentes, finalmente se decide utilizar casi en exclusiva los datos del Instituto Nacional de Estadística (INE), puesto que es la mayor fuente de datos rigurosos disponibles para España. Además, trabajar únicamente con la base de datos del INE asegura una coherencia en las cifras utilizadas, evitando que se pueda dar una discordancia entre estadísticas obtenidas de fuentes distintas al trabajar sobre una misma base de datos. El problema de trabajar con los datos del INE surge de que, por temas de protección de datos, el instituto de estadística español ofrece los datos a nivel de comunidad autónoma y en caso de trabajar con los microdatos por tamaño de ciudad, pero aun así, los datos recogidos con los que se va a trabajar no son exclusivos del municipio de Leganés.

Obtención de datos del INE

Obtener datos del INE en un inicio puede ser una tarea complicada, dado que la página no es muy intuitiva y existe una gran cantidad de encuestas distintas, lo que hace difícil encontrar datos específicos en una primera búsqueda. Por ello, se debe tener claro que datos se necesitan, para luego buscar en qué encuestas se encuentran. Principalmente se ha usado la encuesta continua de hogares (ECH), aunque también se han extraído datos de la encuesta de población activa (EPA).

Dentro de cada encuesta, en el apartado de resultados se tiene acceso a una serie de tablas y gráficos con unas consultas prefijadas por el INE. A estas tablas se les pueden aplicar una serie de filtros que permiten configurarlas para que aparezcan en ellas unos valores u otros. Gracias a estas tablas se pudieron obtener los datos más generales, con los cuales se inició el desarrollo del Familiador. Sin embargo, según se fue diseñando el programa, el nivel de detalle requerido fue aumentando, pues, una aproximación mayor a la realidad requiere de una cantidad mucho mayor de datos específicos que irán aumentando exponencialmente, haciendo que las tablas y gráficos predefinidos del INE dejen de ser útiles, al no contener las estadísticas necesarias para poder seguir con el desarrollo del Familiador. Por esta razón, se tuvieron que buscar otros métodos para ampliar los datos estadísticos. Profundizando en la página del INE se encontró que aparte de las tablas predefinidas, también están disponibles para descargar unos microdatos. Estos microdatos no son completos, puesto que no contienen a toda la población (protección de datos), guardándose para ellos algunos datos como el municipio exacto en el que viven las personas o los nombres de estos.

Procesado de los microdatos del INE

Los microdatos descargados de la ECH vienen divididos en dos archivos, uno en el que se incluyen los datos referentes a las viviendas, y otro con los datos de las personas. En ambos archivos existe una columna con un código identificador de las viviendas, lo que permite que se puedan fusionar los datos de ambos archivos. El INE también incluye con la descarga de los microdatos un programa diseñado en R, orientado para el entorno de desarrollo R Studio y que lee los microdatos y los almacena en dataframes. De este modo se pueden realizar las consultas para obtener las estadísticas necesarias, partiendo de los datos en bruto. Una vez cargados los datos en RStudio, existen multitud de bibliotecas para trabajar con los dataframes como si fuera

una base de datos, para este proyecto se utilizó sqldf. Con ello se podrán realizar joins entre ambas tablas, eliminar filas, hacer media de los valores, etc.

La complejidad de algunas de las consultas ha sido muy elevada, con algunas consultas que han sobrepasado las 200 líneas de código en casos en los que se necesitaban datos muy específicos. Es por esta razón, por la que esta parte del primer prototipo del programa no fue terminada, dejando abierta esta parte para la segunda versión de The Synthetic Population en la que se esperan agregar más estadísticas.

A modo de información adicional, estos microdatos se pueden descargar para el año que el usuario desee, además independientemente del año, existe una variable que indica el tamaño de municipio al que pertenece cada ciudadano, por lo que trabajar con los microdatos añade una precisión que no se podía obtener con las tablas predeterminadas, dado que las consultas aquí se pueden realizar a nivel de comunidad autónoma o a nivel nacional. Por otro lado, si se descargan los microdatos de los años en los que se realiza el censo de población, que se realiza cada 10 años, la cantidad de datos será mucho mayor, dado que existen más variables que no están disponibles para el resto de años. Por ejemplo, se encuentran datos a nivel de sección censal para algunas consultas. Por lo que, si se trabaja con los censos completos, se podría trabajar a nivel de barrio. Actualmente en la página del INE se pueden encontrar los censos completos de los años 2001 y 2011, como ya se ha comentado, el censo de 2021 a fecha de creación de este trabajo no se encontraba todavía disponible.

Generador de tablas a medida del INE

En caso de no saber programar y, por tanto, no poder obtener las estadísticas necesarias mediante los microdatos, el INE proporciona una herramienta muy interesante, pero a la vez compleja. El generador de tablas a medida, esta herramienta que solo está disponible para los años en los que se realiza el censo de población y permite a los usuarios generar sus propias tablas, eligiendo entre una gran cantidad de variables, puesto que estos censos completos son mucho más detallados.

The screenshot shows the 'Generador de tablas a medida del INE' interface. At the top, there are tabs: 'Consulta de Tablas' (selected), 'Visor de Mapas', and 'Área Personal'. Below the tabs is a navigation menu with four steps: 'Paso 1 ¿Qué desea hacer?', 'Paso 2 Seleccione ámbito geográfico', 'Paso 3 Seleccione colectivo', and 'Paso 4 Cree la tabla y establezca filtros'. To the right of the menu are buttons for 'Crear tablas', 'Ayuda', 'Glosario', and links to 'Nacional', 'CC.AA.', 'Provincia', 'Municipio', 'Inframunicipal', 'Personas', 'Viviendas', 'Hogares', and 'Parejas y otros núcleos familiares'. A 'Diseñar tabla' button is also present.

In the center, there is a summary box for 'Ámbito Geográfico' with fields: 'Provincias (Provincia de residencia=[[28 Madrid...]] Ver detalle)', 'Colectivo: Hogares', and 'Filtros: [[2011]]'.

On the left, under 'Seleccionar variables:', there is a tree view of data categories: 'Hogares' (selected), 'Localización', 'Datos del hogar' (selected), 'Composición del hogar' (selected), 'Estructura del hogar', 'Nº de familias en el hogar', 'Nº de núcleos en el hogar', and 'Tipo de hogar' (highlighted). To the right of the tree are two boxes: 'Filas:' containing 'Tipo de hogar' and 'Columnas:' containing 'Tamaño del hogar'. Below these are buttons for 'Ver Tabla' and 'Aplicar e Ir a Filtros'.

At the bottom left, under 'Seleccionar unidad de medida:', there is a dropdown menu with options: 'Número de Hogares' (selected), 'Superficie media por ocupante', 'Tamaño medio del hogar', and '% respecto de la fila'.

Figura 48. Selección de variables en el generador de tablas a medida [70]

Consulta de Tablas Visor de Mapas Área Personal

Paso 1	¿Qué desea hacer?	Crear tablas	Ayuda	Glosario
Paso 2	Seleccione ámbito geográfico	Nacional	CC.AA.	Provincia
Paso 3	Seleccione colectivo	Personas	Viviendas	Hogares

Ámbito Geográfico: Provincias
(Provincia de residencia=[[28 Madrid...]] Ver detalle)

Colectivo: Hogares

Filtros: [[2011]]

Seleccionar variables:

- Hogares
 - Localización
 - Tamaño de municipio de residencia
 - Datos del hogar
 - Datos de la vivienda
 - Datos del edificio

Filtrado actual: Reordenar valores

- No consta
- Menos de 101 habitantes
- De 101 a 500 habitantes
- De 501 a 1.000 habitantes
- De 1.001 a 2.000 habitantes
- De 2.001 a 5.000 habitantes
- De 5.001 a 10.000 habitantes
- De 10.001 a 20.000 habitantes
- De 20.001 a 50.000 habitantes

Filtrado actual:

Los filtros son condiciones que debe satisfacer la información presentada en la tabla (Un filtro: Sexo=Mujer hará que la información de la consulta seleccionada sólo se refiera a mujeres).

Ver Tabla Guardar filtro en área personal Aplicar e ir a diseñar tabla

Figura 49. Selección de filtros en el generador de tablas a medida [70]

Para generar las tablas en esta herramienta hay que seguir cuatro pasos:

1. En este primer paso simplemente se debe seleccionar “Crear tabla”.
2. En el segundo paso se debe seleccionar el ámbito geográfico sobre el que se desea realizar la consulta. Como el creador de tablas a medida usa los censos completos, se pueden crear tablas con datos desde un nivel nacional hasta el nivel inframunicipal. En este último nivel se puede escoger entre distritos o secciones censales, aunque cabe decir que los datos a este nivel son limitados por cuestiones de protección de datos.
3. En este paso se eligen las variables que va a tener la tabla final, como se puede observar en la Figura 48, se dispone de cuatro sistemas de archivos para las categorías Personas, Vivienda, Hogar y Parejas y otros núcleos familiares.
4. En el último paso se crea la tabla como tal, seleccionando las variables y colocándolas en las filas o en las columnas según lo deseé el usuario. Una vez seleccionadas las variables, se puede ver directamente la tabla, o se puede añadir filtros antes de visionar la tabla, como se muestra en la Figura 49.

No obstante, esta herramienta ha sido una fuente secundaria de datos en este trabajo, ya que los datos a los que se podría acceder eran de 2011 y a que las estadísticas necesarias para el trabajo se podían obtener mediante los microdatos en RStudio y, por tanto, no fue necesario sacarle todo el rendimiento potencial al generador de tablas. Cabe comentar, que, mediante la programación de consultas en los microdatos se pueden sacar mejores resultados, pues, el programa generador de tablas no deja generar cualquier tabla, hay que saber escoger las variables y a veces falla. A modo de conclusión se puede decir que la mejor opción combinaria ambos métodos, pero primando a las consultas en R Studio.

A.1.3. Datos de los inmuebles del Catastro

Durante la simulación Matsim calcula los desplazamientos de los agentes por el mapa punto a punto, por ello, necesita asignar los domicilio o los lugares de trabajo a un punto exacto del mapa. Para ello, Matsim admite dos formas de localizar los inmuebles, la primera consiste en asignar manualmente a estas instalaciones un nodo próximo de la red, la segunda consiste en proporcionar las coordenadas exactas de estos edificios, para que luego Matsim asigne automáticamente a cada edificio el nodo que se encuentre más cercano a las coordenadas introducidas, dado que todos los nodos contienen también sus propias coordenadas.

Fuentes iniciales de datos

Las familias sintéticas generadas en el Familiador necesitan una vivienda desde la cual iniciar y terminar su actividad diaria y unos inmuebles de destino para realizar sus actividades. Aunque al inicio de este proyecto se pensó en trabajar con los datos que proporciona el OpenStreetMap, pues estos contenían información de la mayoría de los edificios. Posteriormente, al examinar detenidamente la sede electrónica del catastro, se detecta la existencia de contenido descargable, con datos y con la cartografía de los municipios. Al disponer de estos datos oficiales del estado más rigurosos que los proporcionados por OSM, que provienen de usuarios anónimos que colaboran desinteresadamente, a la vez que se encontraron en la sede del Catastro unos manuales que facilitaban la descarga y el tratamiento de sus datos, se opta por usar el catastro como fuente de datos para asignar viviendas a las familias sintéticas generadas.

Obtención de datos del Catastro

Para descargar los datos del catastro, hay que buscar dentro del apartado de difusión de datos catastrales “descarga de información alfanumérica (formato CAT)”. Al ser datos protegidos, es necesario identificarse mediante un certificado digital o mediante la Cl@ve, una vez identificado, se selecciona la provincia y el municipio del que se desea la información y se descargan los datos. Para poder acceder y trabajar con los datos, además de bajarse este archivo es necesario descargar unas plantillas de Excel que proporciona el Catastro.

Procesado de los microdatos del Catastro

Una vez descargado el archivo y las plantillas de Excel, se puede observar que hay varias plantillas, cada una de estas ofrece distintos datos respectivos a los inmuebles del municipio descargado. En el apartado “ayuda descarga de información alfanumérica” de la página web del catastro, hay varios documentos que muestran el contenido de cada plantilla, para este trabajo se utilizaron dos de ellas:

- **Plantilla tipo 15:** De esta plantilla se han obtenido datos de los bienes inmuebles, como los metros cuadrados de las viviendas.
- **Plantilla tipo 11:** De esta plantilla se obtuvo la localización en coordenadas de las parcelas, que se asociara posteriormente a las viviendas, dado que las coordenadas de las viviendas no aparecen en la plantilla tipo 15.

Para usar las plantillas, hay que renombrar el archivo de los datos del municipio con el nombre “Registro_CAT_tipox” donde x es el número de plantilla que se quiere utilizar, además, ambos archivos tanto la plantilla como los datos, deben estar en la misma carpeta. Al abrir la plantilla, Excel puede pedir que se actualicen los vínculos, para continuar, simplemente habrá que aceptar. Aunque puede que esto no ocurra y que Excel pida directamente que importe el archivo con los datos.

Una vez abierto el archivo, se muestran todos los datos de todas las plantillas, por lo que hay que filtrar por la primera columna “1_tipo_reg”, para dejar solamente los datos de la plantilla deseada. Como el catastro ofrece todos los inmuebles existentes independientemente del uso que se les dé, para seleccionar los inmuebles que atienden a un determinado uso, existe la columna “428_grbice/codusu” por la que se puede filtrar para que solo aparezcan los bienes inmuebles que tengan el uso deseado. Los usos clasificados por el catastro aparecen reflejados en la Figura 50.

CODIGO	USO
A	Almacén -Estacionamiento
V	Residencial
I	Industrial
O	Oficinas
C	Comercial
K	Deportivo
T	Espectáculos
G	Ocio y Hostelería
Y	Sanidad y Beneficencia
E	Cultural
R	Religioso
M	Obras de urbanización y jardinería, suelos sin edificar
P	Edificio singular
B	Almacén agrario
J	Industrial agrario
Z	Agrario

Los datos del catastro fueron necesarios en dos ocasiones, una primera vez para obtener las viviendas de las familias en el módulo del Familiador, por lo que se filtró la columna del uso para que solo aparecieran las viviendas, y en una segunda ocasión para obtener las coordenadas de los destinos de los planes en el Planificador, filtrando la columna de usos para dejar los inmuebles en los que se pudieran desarrollar alguna actividad que no fueran de uso residencial, dejando los siguientes usos: industrial, oficinas, sanidad y beneficencia, deportivo, espectáculos, ocio y hostelería, religioso, comercial y cultural.

Figura 50. Tabla de los usos de los bienes inmuebles [71]

Obtenidos los edificios necesarios, ya solo resta borrar las columnas que no se deseen utilizar, quedándose solo con las pertinentes. Para este trabajo se han utilizado las siguientes columnas:

- Código de la parcela identificativo
- Distrito municipal en el que se sitúa el inmueble
- Metros cuadrados del inmueble
- Coordenada x en coordenadas UTM zona 30N (EPSG:23030)
- Coordenada y en coordenadas UTM zona 30N (EPSG:23030)

Como se comentaba anteriormente, las columnas deseadas se encontraban en dos archivos diferentes debido a las plantillas. Pero dado que ambos archivos contenían la columna código de parcela, se podían unir por esa clave común. Para fusionar ambos archivos en un solo, se hace uso de RStudio y la librería sqldf, importando en dataframes ambos Excel y haciendo un join para obtener un solo archivo con todas las columnas. La tabla resultante de la unión se dividió en 7 archivos, uno por cada distrito, dejando en ambas ocasiones 3 columnas, dos para las coordenadas y una para los metros cuadrados. Al igual que con el censo, los archivos generados fueron guardados en formato csv para que el sintetizador de población aquí desarrollado pudiera leerlos. En cuanto a las coordenadas, dado que estas venían en el formato deseado no hacía falta modificarlas.

A.1.4. Datos de los desplazamientos de la encuesta de movilidad del CRTM

Para crear los planes diarios de las familias sintéticas, en la mayoría de los sintetizadores de población se parte de una encuesta de movilidad que se utiliza como base para poder crear planes sintéticos.

Obtención de los datos

Buscando fuentes de datos para los viajes, se encontró una encuesta de movilidad llevada a cabo por el Consorcio Regional de Transportes de Madrid (CRTM) [55], una empresa pública de

transportes que opera en la comunidad de Madrid e inmediaciones. En la página web¹³ del CRTM estaban disponibles los datos de la encuesta de movilidad listos para la descarga.

Procesado de los microdatos del Catastro

El preprocesamiento de estos datos tenía como objetivo principal la subdivisión del conjunto de familias en distintas tipologías. Sin embargo, como los datos de la encuesta no estaban completos, se tuvieron que realizar distintas transformaciones a los datos antes de que estos pudieran ser utilizados en el Planificador.

Para poder asignar planes reales a las familias sintéticas en el módulo del Planificador, se necesitaban encontrar familias reales similares a las sintéticas. Para conseguir esto, se establecieron unas tipologías de familias en base a una serie de parámetros, por los que ambos tipos de familias podían ser divididos. La clasificación de las familias reales se realizó por medio de una serie de consultas en Python y en R, obteniendo como resultado las 8 tipologías divididas en 8 archivos distintos.

Desafortunadamente la encuesta de movilidad no estaba completa, ya que muchas personas solo contaban con un viaje diario, que muchas veces consistía en un viaje al trabajo por la mañana del que nunca volvían a casa. Por ello, para que la simulación fuera más realista, a todas estas personas se les agrego un viaje de vuelta a su vivienda, para que así tuvieran como mínimo dos viajes diarios. Por otra parte, algunas personas presentaban incoherencias en su planificación diría, ya que les faltaba algún plan intermedio que hacía de nexo entre dos planes. Por ello se ha ilustrado un ejemplo en la Tabla 40, donde se presenta a una persona que tenía un primer plan donde se desplazaba al trabajo desde casa y tenía un segundo plan en el que volvía a su domicilio, pero que tenía como origen un supermercado en vez de su lugar de trabajo. Como no se tenían datos del plan intermedio que hacía de nexo entre los dos planes, se tomó la decisión de modificar los orígenes de los segundos planes para que coincidieran con el destino del primer plan y así conseguir que todas las planificaciones diarias tuvieran coherencia.

Cambio de orígenes para mantener la coherencia				
Acción	Origen 1	Destino 1	Origen 2	Destino 2
-	Domicilio	Trabajo	Supermercado	Domicilio
Cambio origen 2	Domicilio	Trabajo	Trabajo	Domicilio

Tabla 40. Cambio de orígenes

Algunas familias no estaban completas, esto se advirtió al ver que había familias compuestas únicamente por niños, las cuales tuvieron que ser eliminadas. Por otro parte, muchas familias incluían personas sin plan alguno, esto era debido a que estas personas pasaban el día entero en su domicilio y aunque no tenían planes, estas personas eran necesarias para entender la composición de las familiar, por lo que no fueron eliminadas.

Las personas encuestadas del municipio de Leganés eran 2407 de un total de 85065 encuestados, dado que la población de Leganés ascendía a 192.096 personas en 2019, año para el que se han tomado los datos, la opción de coger únicamente los datos de los viajes de los leganenses no era factible. Como la encuesta realizada por el consorcio de transportes tenía datos de toda la comunidad de Madrid, se decidió tomar datos de otros municipios para tener más datos de desplazamientos reales con los que generar la población sintética. Dado que la movilidad de pequeños pueblos o de Madrid capital no tienen ningún parecido con la ciudad de

¹³ Apartado de la página web del [CRTM](#) con los datos de la encuesta

Leganés, que se podría clasificar como una “ciudad dormitorio” de clase media, se decidieron tomar otras ciudades del sur de Madrid con una situación y composición parecidas a las de Leganés.

Las ciudades elegidas fueron Getafe, que contenía 2237 personas encuestados, Alcorcón con 2219 encuestados y por último Móstoles, que aportaba otros 2631 encuestados más. Si se suman los encuestados de los 4 municipios se llega a la cifra de 9394, que supone un 11,04% de los encuestados, tomando a todas estas personas como leganenses, sobre un total de 192.096 ciudadanos, se tienen planes reales para un 4,89% de ciudadanos. Una cifra no muy alta pero que se esperaba suficiente para conseguir unos buenos resultados.

El preprocesado para estos nuevos ciudadanos de Leganés, el cual se encuentra reflejado en la Tabla 41, pasó por cambiar los códigos de origen y destino de los planes que realizaban estas personas dentro de su antiguo municipio, por el código de la ciudad de Leganés. En el caso de que estas personas realizaran planes fuera de su municipio, se mantenían los códigos de las otras ciudades destino, siempre y cuando la ciudad a la que se dirigían no fuera Leganés, dado que, al cambiar la ubicación del hogar de estas personas a Leganés, un viaje por el que una persona salía fuera del municipio, como por ejemplo, un mostolense que viajaba a Leganés, pasaba a viajar de Leganés a Leganés, siendo este desplazamiento un viaje intramunicipal. Para arreglar esto, en estos planes se invirtieron los códigos de municipio, es decir, si el viaje tenía como origen Móstoles y como destino Leganés, esto se invertía para tener como origen la ciudad de Leganés y como destino la ciudad de Móstoles.

Cambio de municipio para encuestados de otras poblaciones						
Persona	Acción	Ciudad en la que vive	Origen 1	Destino 1	Origen 2	Destino 2
1	-	Móstoles	Móstoles	Móstoles	Móstoles	Móstoles
	Cambio de Municipio	Leganés	Leganés	Leganés	Leganés	Leganés
2	-	Móstoles	Móstoles	Madrid	Madrid	Móstoles
	Cambio de Municipio	Leganés	Leganés	Madrid	Madrid	Leganés
3	-	Móstoles	Móstoles	Leganés	Leganés	Móstoles
	Intercambio municipios	Leganés	Leganés	Móstoles	Móstoles	Leganés

Tabla 41. Cambio municipio

Anexo B

B.1. Utilización del software

Para poder utilizar The Synthetic Population se tendrán que obtener todas las estadísticas de hogares utilizadas en el módulo del Familiador para la región que se quiera simular:

- Porcentajes de los hogares por número de habitantes
- Porcentajes de aparición de las tipologías de hogar para cada tamaño de hogar (1, 2, 3,4,5,6,7 o más personas)
- Porcentajes de hogares unifamiliares por edad
- Porcentajes de hogares unifamiliares por sexo para cada intervalo de edad
- Porcentajes de familias monoparentales (madre) por edad de la madre para hogares de 2,3 y 4 personas
- Porcentajes de familias monoparentales (padre) por edad del padre para hogares de 2,3 y 4 personas
- Porcentajes de parejas según la edad de ella
- Diferencia de edad en la pareja
- Porcentajes de parejas heterosexuales y homosexuales masculinas y femeninas
- Porcentajes de hogares compuestos por 2 hermanos según edad
- Porcentajes de hogares compuestos por 2 hermanos según género (2 chicos, 2 chicas o mixto)
- Diferencia de edad entre hermanos
- Diferencia de edad entre hijos
- Porcentajes de hogares de abuelo con nieto por edad del nieto
- Porcentajes de hogares de abuelo con nieto por edad del abuelo
- Porcentajes de hogares de abuelo con nieto por género del nieto
- Porcentajes de hogares de abuelo con nieto por género del abuelo
- Porcentajes de rangos de edad a la que la madre tiene su primer hijo
- Porcentajes de empleo para cada tipología

En caso de realizar el estudio para una región española se recomienda utilizar el INE como fuente principal de datos. En caso de estudiar una región no española, se deberían buscar otras fuentes de datos. Para el caso de no encontrar algunas estadísticas específicas, se podrían suprimir algunas tipologías o simplificar algún proceso de toma de decisiones informado.

Una vez obtenidos estos porcentajes, habrá que introducirlos manualmente en el código, eliminando los datos de Leganés. Una posible mejora para la introducción de los datos en el segundo prototipo, sería la creación de una pequeña interfaz donde introducir los nuevos datos estadísticos o pedir al usuario que incluya todas estas estadísticas en un csv. Con cualquiera de los dos métodos, el programa sobreseguiría donde correspondiese, los porcentajes originales con los nuevos datos. Por otro lado, otra mejora posible a implementar consistiría en la desactivación de algunas tipologías cuando no se disponga de los datos necesarios, generando la población sin estos tipos de hogares.

Para incluir los csv del censo, del catastro y de la encuesta de movilidad, los archivos se tendrán que introducir en sus carpetas correspondientes denominadas Censos, Catastro y Consorcio que se encuentran ubicadas dentro del mismo directorio del programa.

Estructura de csv del censo

El censo tendrá en cada fila el número de personas para cada edad, en la primera fila los recién nacidos, en la segunda las personas con un 1 año, etc., las filas contienen dos columnas, la primera con los hombres y la segunda con las mujeres. Como el programa funciona con distritos, se incluirán tantos archivos como distritos haya.

Edad	Número hombres	Numero mujeres
------	----------------	----------------

Tabla 42. XML censo

Estructura del catastro

El catastro contendrá un archivo con todos los inmuebles de uso no residencial, que utilizará el Planificador y tantos archivos de viviendas, como distritos tenga el municipio, que utilizará el Familiador. Todos los archivos contendrán en cada fila un inmueble y por cada fila se tendrán tres columnas, dos para las coordenadas y una para los metros cuadrados.

Inmueble	Coordenada X	Coordenada Y	m ²
----------	--------------	--------------	----------------

Tabla 43. XML catastro

Estructura de la encuesta de movilidad

Para la encuesta del consorcio se han creado 8 csvs con la misma estructura, uno por cada tipología, cada fila del archivo contiene un plan con 18 columnas de variables. Como el número de variables es muy alto, en la siguiente tabla la fila que contiene todas las variables de cada plan se ha representado con dos filas.

Plan	Id hogar	Id persona	Id viaje	Género	Situación laboral	Permiso de conducir	Hora inicio del plan	Hora fin del plan	Motivo origen
	Motivo destino	Tipo de vehículo	Distancia recorrida	Edad	Número de vehículos en el hogar	Número de personas en el hogar	Número de adultos en el hogar	Código ciudad destino	Código ciudad origen

Tabla 44. XML CRTM

Anexo C

This annex includes a summary of the project and principal ideas.

C.1. Introducción

The information age is generating big changes among society, more than ever with the coronavirus crisis. Digitalization has made an appearance in all fields, many of which were not ready, but the need to continue working has accelerated this transition. These changes are modifying the way people moves and interacts. This movement, constantly accelerated, has a particular impact on cities, which are becoming densely populated, and is altering the coexistence and behavior of their inhabitants.

The need to find solutions to the urban problems is pushing the creation of new technology tools. One of them, perhaps the most important, to study changes in urban dynamics and associated problems is simulation, which allows testing possible improvements to management in multiple fields.

The main objective of this work is to lay the foundations for complex urban transportation simulations, preparing and configuring an agent-based simulation environment in a way that guarantees a simulation representative of reality, facilitating the work for future users.

It should be noted that, to be able to simulate, this type of work requires an arduous task of locating sources and organizing data, instruments, and suitable processes. Creating a map on which to simulate, generating a realistic synthetic population that moves around the map and preparing an environment that allows to simulate, are some of the tasks to be undertaken prior to the simulation. The whole process is approached in a descriptive way to facilitate and enhance its use.

The second purpose of the project is to develop a new method for the generation of synthetic populations based on sociological data, constituting a different approach, focused on studying the basic social factors and structures that define the composition, behavior, and movements of the population. To achieve this, a procedure for the creation of synthetic people based on households is addressed. This approach aims to open a new path in the generation of the initial population for the simulation.

The paper has been structured in 4 chapters:

- **Chapter 1:** Introduction to the project, which covers the principal objectives of it, and motivation that determined its development.
- **Chapter2:** State of art, presents the current situation on which the works starts from, developing aspects related to population, the cities, the modeling based on agents and the chosen simulator, MatSim.
- **Chapter 3:** Solution designing, presents the needs prior to the simulation, developing a solution through software design.
- **Chapter 4:** Evaluation of the developed program, comparing the output with real data

- **Chapter 5:** Conclusions and future development. It includes all the final reflections obtained after the completion of this final thesis and includes the lines of future development that derive from it.

C.2. State of art

C.2.1. City congestion

Cities around the world are becoming increasingly congested due to multiple factors, especially due to overpopulation and migration from rural areas to cities.

This problem, depending on the continent and the degree of development of the countries, will have specific characteristics. In developed countries [4]. The birth rate has decreased drastically [5] leading to an aging population. This phenomenon is particularly significant in Europe, and both Oceania and America are in a similar situation. Other developed economies such as Japan are also affected by this phenomenon. The population of Asia, the world's most populated continent, continues to grow thanks to demographic inertia, although this growth is expected to stagnate around 2060. Last, there is the situation in Africa, a continent that is experiencing a population explosion. Africa has the highest birth rate and expectations indicate that by 2050 the African population will have doubled [9].

The world population is about to reach 8 billion people and is getting closer to the 10 billion mark. Shortages of food and other resources are predicting a possible collapse of the world economy.

The other relevant phenomenon observed in the world's population is its trend toward de-ruralization. It is expected that by 2050, 68% [13] of the population will live in cities. As countries grow economically, their rural populations are concentrating in cities. More and more people are leaving the countryside for urban centers in search of a better future.

In order to stop and/or reverse this trend, some initiatives are being carried out to promote the creation and improvement of services in depopulating areas. Digitalization and the implementation of 5G, the promotion of homeworking or the containment of housing prices in agricultural areas can facilitate this migratory reversion to the countryside.

However, currently, the increase in the population density of cities, and their progressive congestion, is unstoppable. As a result, cities will have to adopt studies, initiatives and changes and their citizens will have to assume modifications in their living habits.

In order to find solutions to the problems that are arising from the increase in the population of cities, many large metropolises are being digitized, giving rise to the concept of Smart Cities [15], which includes everything that surrounds the cities, seeking to improve the quality of life of its inhabitants through sustainable development. It uses information and communication technologies (ICT) and big data to collect different types of data to find solutions.

The application of the Smart Cities concept aims to develop improvement processes in fields as diverse as the economy, government, the environment, social cohesion and mobility within

cities. It is in this last point, where this project aims to provide resources and ideas to support the development of this concept [16].

C.2.2. Evolution of transport

The phenomenon of globalization has connected the world, generating a very complex transport chain that particularly affects mobility in cities in its final stretch. Globalization, the demographic growth of cities and the increase in number of trips and changes in consumer shopping habits (home delivery) are factors that generate a saturation of urban infrastructures. The expansion of e-commerce and its distribution has been a challenge for these companies, which, seeking to reduce costs and delivery times, have invested large resources in achieving maximum efficiency in delivery. These studies have enabled them to acquire knowledge of strategic, tactical and IT reorganization of the movement of goods, which is progressively being applied to the mobility of passengers.

To meet all these new needs, new types of transportation have emerged that overlapped their services with more traditional ones. Offering new services such as carsharing and electric scooters or giving a facelift to existing services such as VTCs, all of which share immediacy, working through apps and opposing collective public transport, generating a greater number of private trips.

On the other hand, new ways of managing travel as a whole have emerged, such as Smart Mobility, which is part of Smart Cities and seeks to achieve cleaner, safer and more efficient mobility, reducing pollution and time wasted in traffic. Smart Mobility encompasses all currently known means of transportation, with the difference that they cease to be the property of the user. In this way, it moves away from the idea of the vehicle as one's own personal transport, and introduces the idea that this vehicle is a service available to people. This form of mobility is known as mobility as a service (MaaS).

This smart mobility is ruled by the principles of [18] **Flexibility**: As there are multiple ways of transport, it is the user who chooses what suits its needs, **Efficiency**: The route offered must be carried in the shortest possible time, avoiding interruptions during the journey, **Integration**: The journey is planned by combining the different means of transport to be used, **use of clean technology**: The different modes of transport used are sought to be as environmentally friendly as possible, focusing on zero emissions and **Safety**: With new technologies and the use of autonomous cars, the aim is to significantly reduce deaths and accidents on journeys, being the autonomous car or intelligent mobility a solution to this problem.

To achieve this improvement, it is essential to advance in the degree of autonomy of vehicles and roads, so that both the use of vehicles and the implementation of decisions and travel plans are made in the most coordinated and cost-effective way possible.

C.2.3. Autonomous car

To classify autonomous cars, the Society of Automotive Engineers of the United States has developed a classification in 6 levels of autonomy, ranging from 0, which lacks driving aid, to 5, which does not give the option of manual driving.

Currently, major brands such as Tesla, Volkswagen or Ford are developing their own Level 4 autonomous systems in which the vehicles will be able to drive by themselves, therefore the

driver will no longer be needed to operate the vehicle, although they still retain the necessary equipment for manual driving, in case they need to assume the responsibility of driving it.

C.2.4. Agents

The agent concept arises from the fusion of various scientific disciplines such as sociology, psychology, software engineering, distributed systems and artificial intelligence. This new discipline begins by addressing problems in a distributed manner and evolves over time. It understands the agent as an isolated computational entity that learns and optimizes its tasks with successive executions.

An agent is a computer system that is capable of executing itself autonomously and flexibly in different situations, being located within an environment in which it must satisfy individual objectives [25].

Agents perceive their environment by sensors. The data obtained is processed by the agent to know the current situation of the environment. With the available information, the agent must decide what actions to take in order to achieve its individual objectives.

The evolution of computer science introduced concurrency and the distribution of resources among systems, giving rise to the distributed artificial intelligence model, which works with more than one system at a time, coordinating them and working semi-autonomously. In this way it is possible to overcome the limitations of working individually.

The application of the advances of distributed artificial intelligence to the field of agents, makes new systems emerge that maintain the idea of autonomy and learning, but It changes from using a single agent to a set of agents working together to achieve a goal.

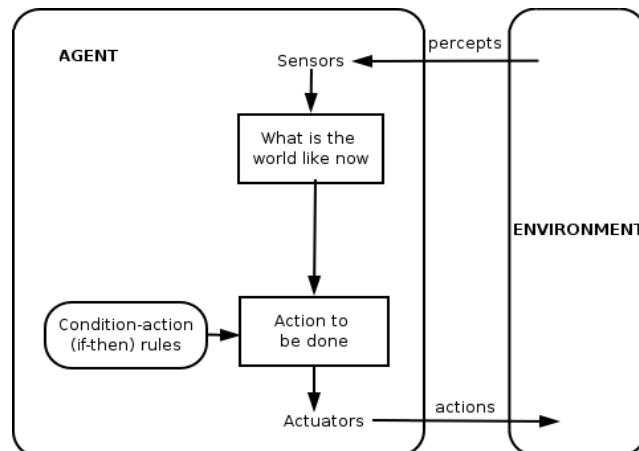


Figure 51. Interaction of agents [26]

C.2.5. Multi-agent systems

Multi-agent systems (MAS) are compounded by a set of agents, which have assigned a mission. They must work together, coordinating, cooperating or negotiating between them in order to fulfill the mission. This task can be divided into independent tasks allowing parallel execution. There is no global control system, data is decentralized and computation is asynchronous. It is the set of agents itself that will assign, by dynamically deciding, the most appropriate tasks to each of them, taking into account the adequacy of an agent to a task and the contribution it makes to the achievement of the common mission. Each agent has an incomplete knowledge.

This limitation can be either about knowledge of the environment, the group's mission, or the intentions of the other agents. The knowledge that each agent has, the possession or not of the required skills and/or its processing capabilities will determine which agent will be the most suitable to perform each of the tasks.

C.2.6. Agent-based model

Agent-Based Modeling (ABM) is a computational model aimed at agent simulation, which arises from multi-agent systems. The difference between ABM and MAS lies in the fact that the Agent-based model aims to understand and explain the reason for the collective behavior of the agents, while MAS is not interested in the behavior that occurs during the simulation, but focuses on the resolution of specific engineering problems, seeking to achieve a result.

This computational model allows the study of the complexity and non-linearity typical of many social phenomena through simulation with agents, being able to recreate and understand how a system works. Therefore, it is often used to study the behavior of natural systems that contain irrationality, instincts, and other behaviors. With it, great advances have been achieved in very diverse fields such as social sciences or biology [33].

C.2.7. MATSIM

Matsim is an open-source framework designed to implement large-scale agent-based transportation simulations. This program arises from the merger of four research streams:

- Microscopic modeling of traffic flow.
- Computational physics.
- Agent-based modeling.
- Coevolutionary algorithms.

Matsim simulates scenarios. A scenario is composed of a population of agents and initial plans, both based on real people and movements that occur in the studied area. These scenarios also include the actual locations of buildings, such as homes or workplaces, as well as the road network. In this scenario the agents will compete with each other to successfully fulfill their plans, modifying their behaviors to reach an equilibrium. Matsim allows the user to configure these scenarios, experimenting with multiple parameters.

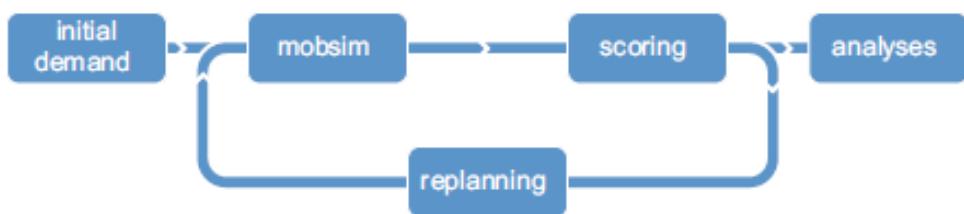


Figure 52. Loop operation of Matsim [35]

Once the system has been configured and the road network and the synthetic population with its daily planning have been introduced, Matsim will perform iterations of a process consisting of the simulation itself, the scoring of the agents after evaluating their actions and the replanning, in which new plans are created from the original daily planning. The system will perform iterations until the configured quantity is reached or until the average score of the population is stabilized, and then the results will be viewed and analyzed.

C.3. Diseño de la solución

Agent-based systems, in order to simulate, need to be provided with data, which are not provided by the existing simulators, i.e. a population and a road network specific to the region to be simulated. For a serious study, the sample populations must be composed of sufficient and representative individuals of the real population to be simulated. To achieve this, the pre-simulation work may require much more work than expected.

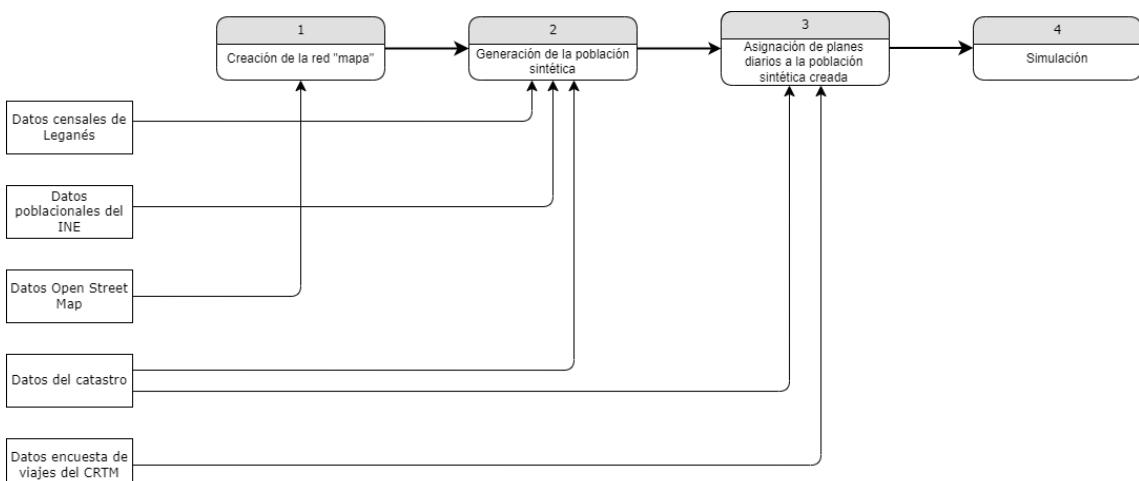


Figure 53. Map of the processes to be implemented in order to be able to simulate

Figure 53 shows a simplified scheme of the development process to be followed in order to be able to simulate with Matsim. The first task consists of creating the road network where the simulation will take place, then generating the synthetic population and finally assigning the daily planning to the synthetic people created.

C.3.1. Network generation

In the first step, the Leganés road network will be generated. For this, the data obtained from OpenStreetMap, a collaborative project that has generated the complete map of the Earth with its buildings, streets, neighborhoods, etc., will be used. The OSM data will be extracted using the JOSM program. This software was chosen because it contains a specific extension for Matsim created by Andreas Neumann and Michael Zilske [37], which facilitates obtaining the necessary data to simulate, by giving as output a file with the structure required by Matsim. Obtaining a road network like the one in Figure 54.



Figure 54. Map of the road Leganés network of obtained with the plugin [41]

C.3.2. Generating the population

The complete real population of a region is not available due to the limitations established by the data protection legislation of many states. Therefore, to generate the complete simulation of population, a number of methods and algorithms have been developed. Which, starting from a small portion of real data and constraints obtained from different sources, generate a realistic synthetic population of the region to be simulated. These programs are called population synthesizers and each one uses different methods to achieve the target population

Given a set of demographic data belonging to a subset of the real population, a population synthesizer seeks to adjust this sample, selecting the most influential variables and adjusting their weights to create a synthetic population from them.

Population synthesis can be considered from a mathematical point of view as an optimization problem, which seeks to minimize the differences between the generated and desired frequencies of the control variables, looking for optimal weights for the households. This optimization can be formulated by means of equations, showing different approximations to the final synthetic population.

There are multiple approaches for the generation of synthetic populations, as reflected by Zhuge, Shao, Wang and Hu [44], which can be grouped into 5 main types:

- Demographic distribution models
- Simulation-based models
- Data Constriction models
- Entropy Optimization Models
- Fitness-based Models

When these approaches were reviewed, it was observed that most of them focus on finding the factors that define the census sample on which they work, using algorithms, and then extrapolate this knowledge to the total population, leaving aside the social structures of the population.

This paper aims to take a different approach, focusing on analyzing the social structures left out by the rest, and then generating the population from the data extracted from these structures.

Although family structures differ depending on the region in which the study is carried out, there are basic structures such as, the family composed of parents and their children. Which are present in most cultures, in different percentages of appearance, but present. However, for the application of this model to other regions, it will be essential to contextualize the structures identified with the social structures typical of the new region, which could include new typologies to be incorporated into the algorithm.

To initiate this approach, prior to the creation of the software, a first social study will be developed. The aim is to identify the basic social structures, the different types of families and their percentage of occurrence for the region under study. This study uses data from the city of Leganés, where the Carlos III University of Madrid is located.

C.3.2.1. The Synthetic Population

The population synthesizer developed in this document is specifically created to be able to simulate in MatSim. However, although the *output* is targeted for MatSim, both the population and the plans generated are valid for any other agent-based simulator by restructuring the data and modifying the output.

It has been named "The Synthetic Population" and is composed of two main modules. The first one, named "The Familiador", is in charge of generating synthetic families from the data extracted from the census, the cadastre and from the national institute of statistics (INE). Which contains the number of Leganés population, the real housing of the municipality and the statistical data of the population. The second module, called " El planificador", uses data extracted from a travel survey [55], which contains data from real trips made by citizens of Madrid and allows assigning plans of real surveyed families to the synthetic families generated by the Familiador.

C.3.2.2. Generation of families, El Familiador

Humans are social beings in constant interaction with other human beings. These interactions are guided by a series of cultural rules defined by society. Although these rules differ according to the culture of each context, they all share the fact that individual mobility is influenced by their closest social nucleus. People plan their daily lives around basic objectives that are affected by those of the people around them. Although there are many individuals who influence the mobility of each one, the greatest influence is determined by the people with whom they live, i.e. the family, the members of their household, whether or not they are related to each other. Therefore, it makes sense to treat cohabitants in a household as "families" when processing their data, even if they are managed differently.

As mobility is strongly influenced by the household, the generation of the synthetic population will be carried out by creating households and not individuals. Prior to the step of programming the generation of households, in order to support the sociological approach developed in this

project, a reflection was carried out to find out what these households are like and to understand how their members move. A series of factors were defined to help characterize their behavior.

The factors analyzed in relation to household composition and mobility were the following: age, level of education, employment status of household members, nationality/culture of origin, per capita income and size of houses.

Not all of the above factors could finally be used in the analysis due to time constraints. Obtaining data for some of them was very complex and it was decided to leave their implementation for a second prototype.

The analysis of the aforementioned factors led to the elaboration of a series of typologies related to household composition. They were classified according to the number and type of members of which each household is composed, as shown in Figura 18 and Figura 19.

Once the household typologies had been identified, statistical data from the INE began to be collected, with which "El Familiador" was developed. The program allows, by randomly choosing people from the census, to generate types of households that will appear with a frequency based on their actual percentages of appearance. Once each family has been created, the program will assign it a real house from the cadastre. This step completes the generation of synthetic families.

C.3.2.3. Generation of plans. El Planificador

Once the synthetic population has been generated, the next step will be to assign to these families a series of initial mobility plans that will be used by the simulation agents to move around the map of Leganés. This initial plan generator, after successive simulations, will optimize them.

As was the case with the creation of the population, all the studies consulted on the generation of daily plans do not study mobility using sociological variables characteristic of the population, but generate the plans based on real plans. If it is wished to study mobility based on an analysis of social structures, it will be necessary to study the factors that shape and condition this mobility. Some of these factors are common to those analyzed in the "Familiador" section, while other factors that only affected mobility are considered only in this section.

Factors such as the date on which mobility occurs (daily or weekend days or even holidays such as Christmas or summer vacations with different habits and cultural loads), the type of street and how this affects the mobility that takes place on it, or the weather (rain makes more people take the car to move around), and a long list of additional factors can be integrated in the future.

Defining mobility, taking into account social structures, is much more complex than defining family composition due to the fact that mobility is influenced by a greater number of factors and, also, due to the scarcity of data that can be obtained from government sources, that collect a smaller number of data related to mobility. Finally, given the complexity of the issue and for practical reasons, it was decided to prioritize the main objective of this work over the search for a new approach to the generation of synthetic populations. Taking a more traditional approach in which, the actual plans of a travel survey are used as the basis for generating synthetic daily planning. To do so, the real plans obtained from the CRTM have been used, eliminating the coordinates of the origin and destination locations while keeping the rest of the attributes (such as the type of transport chosen for the trip, etc.).

In order to be able to assign these plans to the synthetic families created, the program will search for real families that perform a similar daily mobility. The greater the similarity between the two families, the better the results obtained. Both families, real and synthetic, should have the same composition, the same level of education and/or the same employment situation. To implement this, it is necessary to establish household typologies. The initial idea was to use the same household typologies used in "the Familiador", however, when it was put into practice, a series of problems arose:

- The travel survey used did not contain information on the kinship relationship between the cohabitants of the surveyed households, which made it impossible to identify the most specific types of households.
- There was not enough data available to make a classification as exhaustive as the one elaborated in the first module.
- In order to correctly represent mobility, it would have been necessary to include other factors chosen from among those that do not affect household composition but do affect mobility. To do so, would have meant that the multiple typologies developed in the Familiador would have to be subdivided according to these other factors, complicating the process enormously.

In order to overcome these problems, a more general classification was elaborated in which four dichotomous factors were taken into account: household size (single/multi-person), employment status (employed/unemployed), presence of children (yes/no) and whether the family is single-parent (yes/no). From the combination of these, eight household typologies were obtained. Which were used to classify both the synthetic families of the Familiador and the real families of the travel survey:

1. Single Person - Unemployed - No children
2. Single Person - Working - No children
3. Multipersonal - Unemployed - No children
4. Multipersonal - Working - No children
5. Multipersonal - Unemployed - With children - Single parent
6. Multipersonal - Unemployed - With children – Non single parent
7. Multipersonal - Working - With children - Single parent
8. Multipersonal - Working - With children – Non single parent

When the program performs the transfer of plans, it will first match by type each synthetic household with a real one, then check that the number of family members matches. It will perform a member-to-member plan assignment, by similarity of age and gender.

Coordinate calculation: As previously mentioned, the transferred plans did not contain the origin or destination coordinates. In order to calculate these coordinates, a differentiation has been made between three types of trips:

1. **Trips that are made to the home:** In this case as these people had already been assigned a real house from the Cadastre, the real coordinates of it will be used.
2. **Trips to another municipality:** The agents, in this type of trip, leave the map of Leganés. For these trips, only the outbound journey from the municipality would be calculated, assigning a route that would go from the starting point of the previous plan to some

outbound coordinates that will vary depending on the destination city. To each destination municipality with a significant flow of trips, specific departure coordinates were attributed.

3. **Trips within the municipality with a destination different from home:** This type of trip is the most complex to program. You need to calculate the coordinates based on the origin, the distance and the type of activity you want to carry out at the destination. It will take into account the infrastructures of the municipality extracted from data contained in the cadastre and in the travel survey. In order to unify both classifications, 7 types of trips were created and used to classify all the infrastructures of Leganés. All actual plans were attributed, depending on their destination, one of these 7 trip types.

To calculate the new coordinates of the plan, a circle is drawn around the origin point with a radius equal to the distance traveled. This makes it possible to identify all the possible points that could be chosen as a destination. To choose one of the selected points, it is made use of the infrastructure surrounding each of them. To do so, the program has been provided with a zoned map of Leganés, creating a raster where each cell will include a series of properties. As the coordinates of each point are located within a quadrant of the map, it will be possible to calculate the cell in which each one is located, acquiring the list of infrastructures of the cell.

Therefore, to calculate the target coordinates of a plan, all possible target cells will be calculated and to choose among these. The amount of infrastructure of the type of activity to be done in the target will be taken into account. Once a cell is chosen by probabilistic means, the coordinates of the point are obtained and assigned to the plan, completing it.

C.4. Evaluation

In order to check if the synthetic population resulting from the software developed in this work adjusts to the real population, the 8 typologies of households obtained from The Synthetic Population will be compared with the same 8 typologies obtained from the INE microdata, checking if the percentages of one correspond to those of the other.



Figure 55. Calculation of the new destination

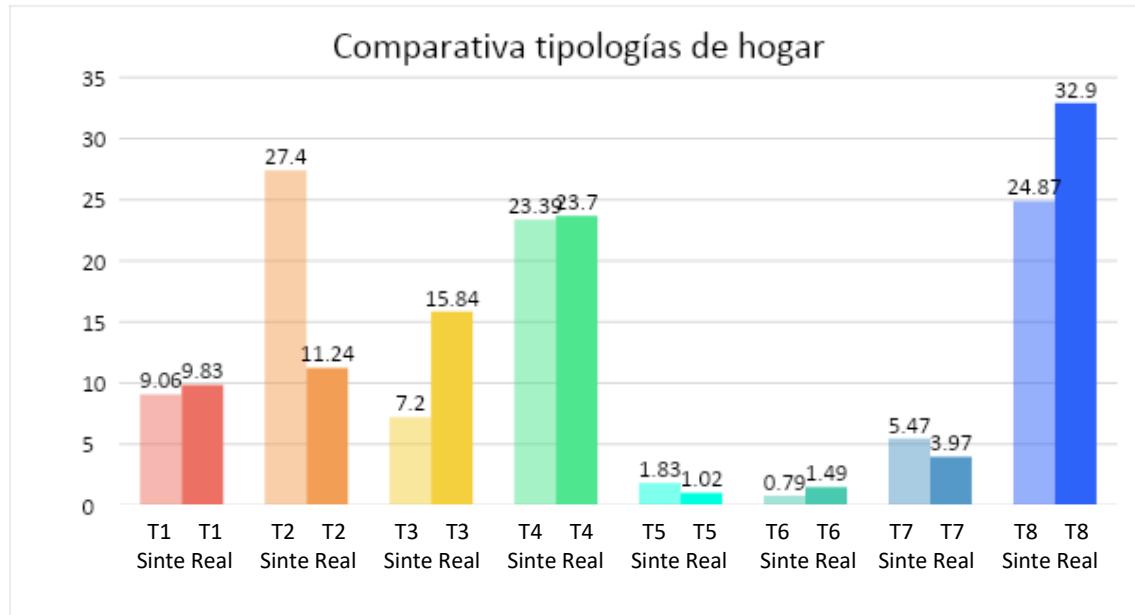


Figure 56. Comparison between real and synthetic home typologies

As can be seen in Figure 56, some household typologies fit better than others. These fit problems are partly due to the fact that the first prototype of the developed software does not contain most of the identified household typologies. The amount of time required for the development of this area would have prevented achieving the main objective of this work, which is why the generation of all typologies was limited.

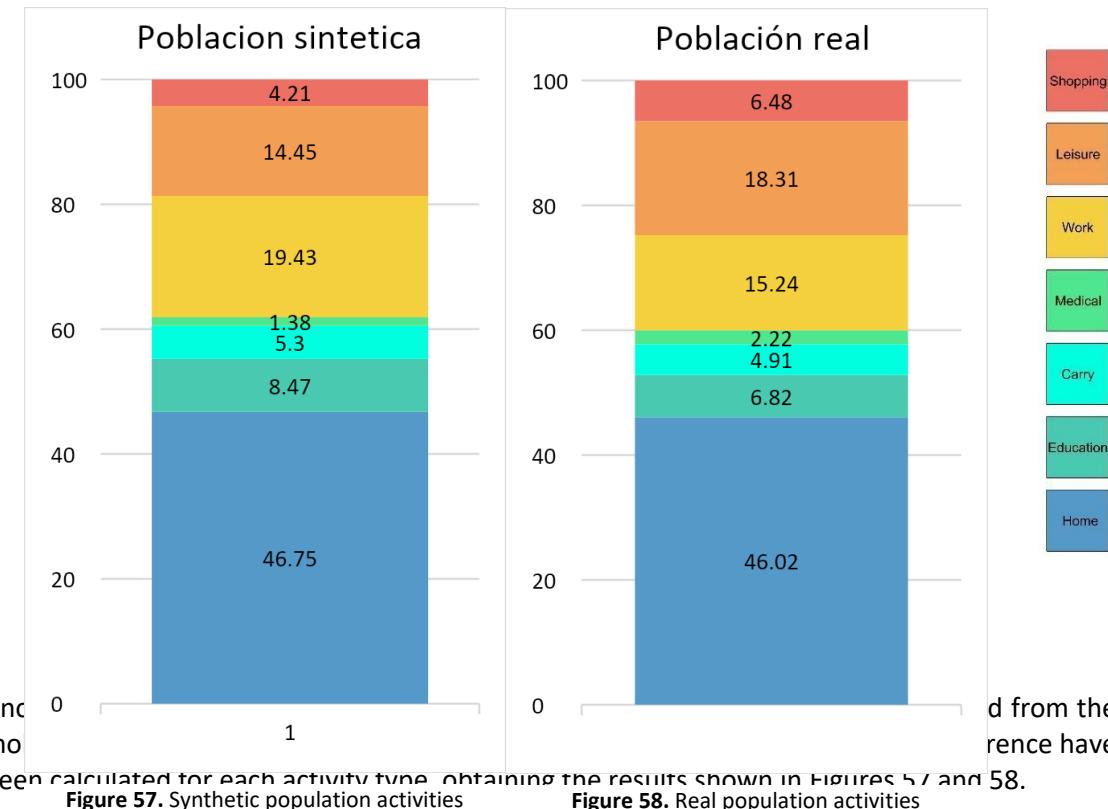


Figure 57. Synthetic population activities

Figure 58. Real population activities

The types of activities performed by the synthetic population are more closely matched to those of the real population. However, this fit is not perfect, partly because the under- or over-representation of some types of households, seen in the previous comparison, leads to the under- or over-representation of some types of activities. Each type of household performs a

greater number of activities of one type, and since not all types of households are represented in an equivalent manner, not all types of activities are accurate.

C.5. Conclusions

The main objective of this project was the preparation and configuration of a simulation environment, exposing and developing all the necessary processes prior to the simulation.

Throughout the different parts of this work, the necessary knowledge prior to simulation has been developed. The basic principles of agents have been defined, then multi-agent systems and agent-based simulation have been described. The agent-based simulator used in this project: MatSim, its basic operation and components, was introduced. The reasons that led to the creation of a population and a road network for simulation purposes were presented.

Subsequently, the different ways used to obtain the necessary files to simulate were developed. The essential data and the way to acquire them were specified. The software developed in this work was also presented, explaining its operation.

Once this point has been reached, it can be stated that the main objective of the work has been achieved.

The second objective of the project was to develop a new method for the generation of synthetic populations based on sociological variables, a different approach, which focuses on studying the factors that define the composition and movements of households.

At the end of this work it has been able to develop a software capable of generating a complete synthetic population of a desired region from statistical data of the target population. The file obtained as output of this software has the correct format and works in MatSim.

The results of the program were evaluated at the previous section where the families created were compared with the actual families and the total trip types assigned were compared with the actual trips from the mobility survey. The percentages in most categories were very similar, although with some exceptions. Given the initial state of development of the program, the results obtained are considered satisfactory and new typologies will continue to be implemented in subsequent prototypes with which the results are expected to improve.

The main difficulty of this work has been the complexity, immediacy and non-linearity typical of many social phenomena. It is a challenge to represent social relationships that are associated with randomness and irrationality, making it difficult to achieve the optimal solution. Taking into account human complexity makes it necessary to process a large number of variables, whose data are not always available, in order to obtain results that conform to reality.

Each time a new phase of a project has been entered, it has been found that the original simple ideas are far from reality, and as each idea is explored in more depth, new fields of research are opened up.