



Nacho Reyes Vázquez: Project Leader

Diego Pitarch Falomir: Project Administrator

Jose Antonio Rodrigo Cortés: Technician Specialist

Guillermo Andrés Roncancio Salamanca: Project Communications Coordinator

Carlos Buenrostro Valverde: Business Specialist

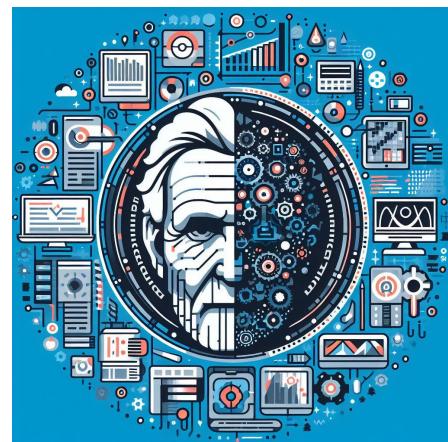
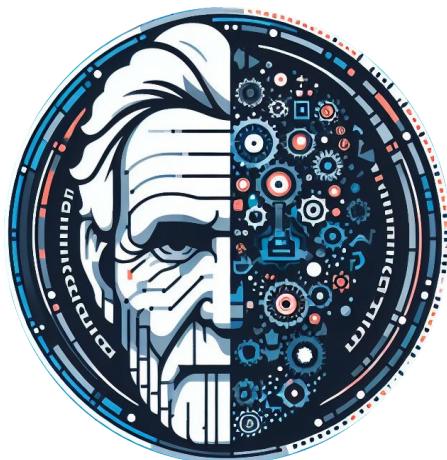


Presentación empresa:

La empresa Golden Analytics, fundada en el año 2023 por Muten Röshi, tiene como objetivo el análisis y mejora de los sistemas implantados en la actualidad. El core de la empresa se basa principalmente en la aplicación de la justicia en la toma de decisiones gracias al análisis de datos.

El caso de estudio presentado es la asignación justa de las plazas hoteleras en los viajes del Imserso para nuestros mayores. Para ello, el uso de las nuevas tecnologías es fundamental, con la modificación de variables y el análisis de los datos. Esto proveerá el mejor método de asignación posible.

Para la creación del logo, se ha decidido aunar los dos pilares fundamentales del caso de estudio presentado: Las nuevas tecnologías y los mayores de nuestro país. Con la aplicación de la Inteligencia Artificial, de la mano de Dall-e e introduciendo el siguiente prompt “Generar logotipo para una empresa de analítica de datos que debe mejorar el sistema de asignación de plazas hoteleras para ancianos del Imserso”, el resultado ha sido el siguiente:





Study Case: Asignación plazas hoteleras Imserso.

Ante una necesidad de modernización por parte del Imserso, estos se han propuesto abrir un concurso público para que el diseño de la plataforma de asignación de plazas hoteleras para los ancianos sea mucho más moderna, pero sobre todo justa.

Para ello, la empresa Golden Analytics ha decidido presentarse con el fin de mostrar su producto y modelo de negocio para hacerse con la plaza. Este proyecto ha pasado por distintas fases que se detallan a continuación:

Análisis preliminar:

En esta primera etapa del proyecto se ha decidido estudiar el actual sistema que utiliza el Imserso para la adjudicación de las plazas hoteleras a los mayores. Para ello se han tenido en cuenta las distintas variables de asignación establecidas, así como el análisis de los distintos puntos de mejora que pueda tener el actual modelo.

Variables:

Una vez realizado el análisis del modelo actual del Imserso, se han añadido nuevas variables, bajo criterio de la compañía, que harán que el modelo de adjudicación que hasta hoy en día existe, sea aún más justo:

- **Edad.** Al igual que el IMSERSO, nos parece coherente favorecer que viajen las personas de mayor edad según el baremo siguiente, que es el mismo que el sistema actual:

- Menos de 60 años: 1 punto.
- 60 años: 2 puntos.
- Entre 61 y 78 años: 2 puntos más 1 punto por cada año adicional.
- 78 años o más: 20 puntos.

- **Años Trabajados.** Se considera que aquellas personas que han desarrollado una larga carrera profesional en el país tengan un extra de bonificación frente a aquellas que no.

- Menos de 20 años: 10 puntos.
- Entre 20 y 25 años: 20 puntos.
- Entre 26 y 30 años: 30 puntos.
- 30 años o más: 40 puntos.



- **Movilidad.** Según el grado de movilidad del individuo (limitada, media o perfecta) se establece un score que favorezca a aquel que no necesite prácticamente ayuda para poder realizar sus actividades diarias.

- Movilidad limitada: 10 puntos.
- Movilidad media: 20 puntos.
- Movilidad perfecta: 30 puntos.

- **Discapacidad.** Según si es discapacitado o no, se le otorgan puntos para favorecer a las personas con discapacidad.

- Sin discapacidad: 0 puntos.
- Con discapacidad: 25 puntos.

- **Enfermedad.** Similar a la variable movilidad, si la persona no sufre de ninguna dolencia o enfermedad, se verá beneficiada en su score final ya que el coste o riesgo para el viaje es inferior frente a la persona que lo sufre. Para ello, nuestra empresa ha subcontratado a un equipo profesional médico encargado de realizar dichas revisiones.

- Sin enfermedad: 50 puntos.
- Con enfermedad: 0 puntos.

- **Familia Numerosa.** Si un individuo, tiene la condición de familia numerosa no pensamos que deba tener puntos adicionales, pero en el caso de tener un hijo dependiente de él, otorgamos puntos si el hijo es menor de 26 años en función de si éste trabaja o no.

- Sin familia numerosa: 0 puntos.
- Familia numerosa con hijo menor de 26 que no trabaja: 50 puntos.
- Familia numerosa con hijo menor de 26 que trabaja: 15 puntos.

- **Habitación.** Se considera esta variable para el score final ya que siempre es más económico poder reservar habitaciones dobles que una individual. Es por ello que, si el anciano escoge la habitación compartida, se verá beneficiado.

- Habitación doble: 20 puntos.
- Habitación individual: 0 puntos.

- **Renta.** Es importante tener en cuenta la variable de la renta ya que es, probablemente, uno de los aspectos más importantes a la hora de seleccionar un candidato. Por ello se establece un score diferente según el grado de ingresos en los que se encuentre.



- Renta < 1: 50 puntos.
- 1 < Renta < 8500: 40 puntos.
- 8500 < Renta < 13000: 30 puntos.
- 13000 < Renta < 16000: 20 puntos.
- 16000 < Renta < 19000: 10 puntos.
- 19000 < Renta < 22000: 0 puntos.

- **Vehículo.** Se considera la variable vehículo dentro del modelo debido a que aquellas personas, o familias que dispongan de un vehículo propio tienen una mayor oportunidad de viajar a lo largo del año que aquellas que no disponen de la titularidad de un vehículo, quienes recibirán una pequeña bonificación.

- Sin vehículo: 20 puntos.
- Con vehículo: 0 puntos.

- **Años Anteriores Viajando.** En esta variable, se evalúan tres factores. Si el individuo está en lista de espera o no, si ha viajado en el año 2021 y/o en el 2022 y cuántos viajes ha realizado para otorgar en cada caso la puntuación correspondiente:

- Si está en lista de espera y viajó el año 2021 y el 2022 una vez cada año, 40 puntos.
- Si está en lista de espera y no viajó los dos últimos años, 70 puntos.
- Si está en lista de espera y no viajó el 2021 pero si el 2022 menos de dos veces, 50 puntos.
- Si está en lista de espera y viajó el 2021 menos de dos veces pero no el 2022, 60 puntos.
- Si está en lista de espera y viajó los dos años anteriores 5 puntos.
- Si está en lista de espera y no viajó el 2021 pero si el 2022 más de una vez, 10 puntos.
- Si está en lista de espera y viajó el 2021 más de una vez pero no el 2022, 15 puntos.
- Si no está en lista de espera, 0 puntos.

- **Tipo de destino.** Según el tipo de destino elegido, otorgamos puntos, beneficiando los destinos que bajo nuestro criterio interesa fomentar.

- Si el tipo de destino es 'circuitos culturales', 50 puntos
- Si el tipo de destino es 'turismo_naturaleza', 35 puntos
- Si el tipo de destino es 'zona baleares', 20 puntos
- Si el tipo de destino es 'capitales provincia', 0 puntos



- Si el tipo de destino es 'zona costera', 10 puntos
- Si el tipo de destino es 'viaje ceuta melilla', 70 puntos
- Si el tipo de destino es 'zona canarias', 30 puntos

- **Precio.** Al igual que la variable anterior, hemos optado por evaluar el tipo de destino elegido favoreciendo unos respecto de otros y si tienen el transporte incluido o no según los siguientes criterios:

- Zona Costera Peninsular con transporte 10 días, 12 puntos
- Zona Costera Peninsular con transporte 8 días, 11 puntos
- Zona Costera Peninsular sin transporte 10 días, 13 puntos
- Zona Costera Peninsular sin transporte 8 días, 5 puntos
- Zona Costera Insular: Illes Balears con transporte 10 días, 8 puntos
- Zona Costera Insular: Illes Balears con transporte 8 días, 9 puntos
- Zona Costera Insular: Illes Balears sin transporte 10 días, 14 puntos
- Zona Costera Insular: Illes Balears sin transporte 8 días, 3 puntos
- Zona Costera Insular: Canarias con transporte 10 días, 2 puntos
- Zona Costera Insular: Canarias con transporte 8 días, 10 puntos
- Zona Costera Insular: Canarias sin transporte 10 días, 15 puntos
- Zona Costera Insular: Canarias sin transporte 8 días, 1 punto.
- Turismo de Escapada: Circuitos culturales 6 días, 4 puntos
- Turismo de Escapada: T. naturaleza 5 días, 7 puntos
- Turismo de Escapada: Capitales de provincia), 16 puntos
- Turismo de Escapada:Ceuta o Melilla , 6 puntos

- **Meses.** Con esta variable pretendemos favorecer los viajes en los que la demanda turística es menor

- Octubre, 0 puntos
- Noviembre, 10 puntos
- Diciembre, 15 puntos
- Enero, 30 puntos
- Febrero, 25 puntos
- Marzo, 20 puntos
- Abril, 10 puntos
- Mayo, 5 puntos
- Junio, 0 puntos



- **Menú.** Definimos esta variable para fomentar el gasto tanto en los hoteles como fuera de ellos, pues en fechas señaladas se ofrecen menús especiales como en Nochevieja y el precio del viaje es mayor.

- Sin solicitud de menú: 0 puntos.
- Con solicitud de menú: 20 puntos.

- **Pensión.** Puesto que el impuesto sobre la renta puede reflejar otras rentas a parte de la pensión que el individuo recibe, hemos decidido también evaluar esta condición respecto a las siguientes condiciones:

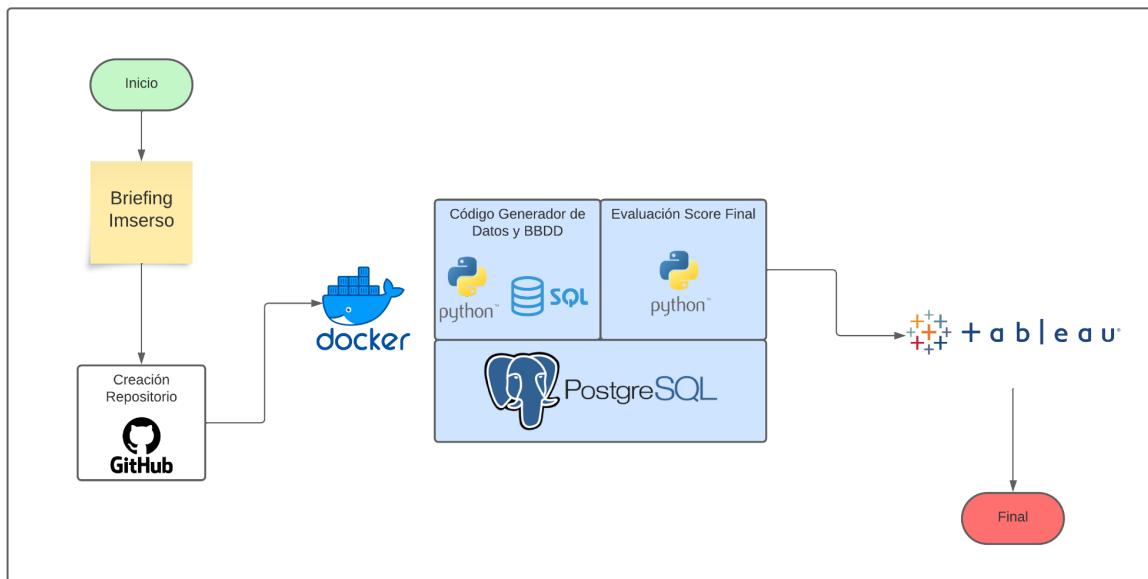
- Si la pensión está entre 13076€ y 6789€, 50 puntos
- Si la pensión está entre 19365€ y 13075€, 40 puntos
- Si la pensión está entre 25654€ y 19364€, 30 puntos
- Si la pensión está entre 31943€ y 25653€, 20 puntos
- Si la pensión es mayor de 31942, 10 puntos

- **Reevaluación Final.** Adicionalmente, para que una persona no pueda elegir varios viajes y que otra no elija ninguno reevaluamos el score final y restamos 80 puntos a las solicitudes de un individuo que haya sido elegido más de una vez, dejando su puntuación máxima intacta. Este criterio intenta favorecer que el máximo número de individuos puedan optar a obtener un viaje frente a que un individuo repita un viaje.

Desarrollo:

Tras elegir las variables que se introducirán en el modelo de datos, se comienza con el desarrollo de producto. Para ello, los departamentos implicados en el proceso se alinean con el objetivo y estrategia marcados.

En primer lugar, una vez dado el brief de la tarea a realizar, se decide establecer el fluograma de trabajo que ilustrará las herramientas y pasos a seguir en cada etapa del proyecto. Para ello, se utilizará la herramienta Lucidchart. Esta, con una interfaz sencilla e intuitiva, permite generar de forma fácil y sencilla el fluograma del Data Project.



Seguido se creó el repositorio privado de GitHub entre los distintos colaboradores del equipo. En él, a través de ramas y organizado en carpetas, se trabajará y se subirá todo el código generado, así como los resultados del mismo.

Una vez creado el repositorio, se pasa a la realización de un Producto Mínimo Viable (PMV), en el que, con lo más sencillo, corriese todo el flujo de trabajo. Reportando así el resultado esperado. Para la realización del mismo se preparó un dockerfile y un docker-compose.yml que levantaría los contenedores, la base de datos en un servidor Postgres y las diversas conexiones a la misma. Además, un pequeño script de python que generaba datos y un init.sql que generaba una tabla en Postgres. Ejecutando todo, se generó el resultado esperado, el PMV.

El siguiente paso que se siguió en el equipo de desarrollo fue la escalabilidad del proyecto. Dividido el trabajo entre los componentes se comenzó por la creación de las variables que influirían en el modelo justo. También, por el desarrollo del código generador de los datos. Se utilizó Python junto a sus amplias bibliotecas como: Pandas, Psycopg2, Faker, Gender-guesser y unicode. El origen de los datos no proviene de ningún dataset establecido. Se han generado de forma aleatoria y con identidades inventadas gracias a la biblioteca Faker de Python.

En paralelo, los miembros encargados de la Base de Datos se encargaron de generar un archivo .sql que creara las tablas en las que los datos iban a volcarse.



Producción y entrega:

En la fase final del proyecto, una vez creado todo el ecosistema en el que iba a correr el producto, se comenzaron las pruebas end to end donde se comprobó que en todas las máquinas de los integrantes del grupo corría y generaba el entregable.

El entregable final consiste en un archivo generado .json que junta todas aquellas personas que han sido seleccionadas para poder realizar un viaje, según las características y puntuación final.

Diseño de Arquitectura

En este apartado vamos a describir la arquitectura de aplicaciones que hemos utilizado para crear este nuevo modelo de scoring por puntos.

Todo nuestro modelo ha sido desarrollado en Visual Studio Code mediante las siguientes aplicaciones:

1. Python para realizar una función y obtener el score, generar datos aleatorios.
2. Docker para gestionar el Docker-compose y poder trabajar fácilmente con varios contenedores, además de escribir el Dockerfile para la construcción de la imagen.
3. SQL para inicializar la base de datos y que los datos se almacenen dentro de ella de la forma correcta.

PostgreSQL para poder almacenar de forma ordenada los datos generados, además de poder realizar búsquedas dentro de la base de datos mediante Dbeaver.

Docker para crear una imagen que contenga todo lo necesario para que al desplegar el contenedor, además, con docker-compose podemos gestionar correctamente varios contenedores a la vez.

La elección de las herramientas que hemos utilizado se debe a que las hemos conocido en clase y pensamos podrían ser de gran ayuda:

1. Python al ser una de las aplicaciones de desarrollo de código con la que estamos más familiarizados, pensamos que era la elección idónea para generar nuestro proyecto. Este lenguaje de programación ofrece ventajas como una amplia biblioteca que nos permite la manipulación de datos, aprovechando funciones preexistentes, facilitando el desarrollo de manera ágil y colaborativa. A partir de Python hemos



logrado crear el dataset con 50.000 datos usando librerías como pandas, numpy y Detector; de igual manera ha facilitado la conexión con PostgreSQL y Docker, vitales en la solución creada y que explicaremos más adelante.

2. Dbeaver es la herramienta que decidimos utilizar para la base de datos, ya que tenemos un mayor conocimiento de ella, adicionalmente la posibilidad de conectarnos a diferentes bases de datos de manera sencilla y trabajar con sistemas de gestión de datos como PostgreSQL nos permite interactuar y desarrollar el proyecto de una forma más eficiente.
3. Docker es la aplicación de contenerización que más confianza nos daba a la hora de desarrollar un contenedor y debemos utilizarla para que al ejecutar el contenedor se despliegue de forma automática tanto los datos como la base de datos y que estos se almacenen en ella.
Además, Docker Compose nos permite personalizar el entorno de desarrollo evitando errores que se deriven de ello, facilitando el trabajo de los integrantes del equipo y garantizando una misma configuración.
4. Tableau como plataforma de visualización nos permite analizar y graficar los datos generados de la solución desarrollada para el Imserso; como producto de esto, será posible mostrar conclusiones y las ventajas de la herramienta creada por nuestro equipo de trabajo.

Para este trabajo hemos considerado trabajar con un dataset generado por nosotros, por la complejidad de encontrar un dataset con datos de personas reales, pensamos además que generarlo por nosotros mismos nos daría un mayor dinamismo ya que podíamos generar variables específicas.

El dataset lo hemos creado otorgando valores aleatorios para las variables que pensamos que sería interesante evaluar. Dentro de las variables hay algunas como el nombre o el sexo, las cuales no repercuten en el score, pero hay otras como la renta, la edad o el mes en el que se realiza el viaje en las cuales si que debemos obtener un score de ellas.



Modelo de datos

El modelo de datos implementado consta de una base de datos relacional compuesta por 12 tablas que almacenan la información generada.

