



## Predicción de demanda de alquileres por países

Repositorio de Github: [https://github.com/pablopuch/sagulpa\\_final.git](https://github.com/pablopuch/sagulpa_final.git)



## 1. Introducción

En este proyecto final del curso de Especialización de Inteligencia Artificial y Big Data del centro I.E.S El Rincón en el año 2022/2023, he realizado un trabajo sobre la predicción de demanda de alquiler de bicicletas de la empresa Sagulpa por países, en el cual data una fecha, lugar y nacionalidad cuántos números de turistas van a alquilar sitycleta.



## HERRAMIENTAS

### **Power Bi**

Power BI es una herramienta de visualización y análisis de datos desarrollada por Microsoft. Permite a los usuarios conectarse a diversas fuentes de datos, combinar información de diferentes orígenes y crear informes interactivos y paneles de control personalizados.

Con Power BI, puedes importar datos desde una amplia variedad de fuentes, como bases de datos, archivos de Excel, servicios en la nube, entre otros. Luego, puedes modelar y transformar los datos según tus necesidades utilizando el lenguaje de fórmulas DAX (Data Analysis Expressions). También puedes realizar cálculos, crear relaciones entre tablas y definir medidas para agregar y analizar los datos.

La cual hemos utilizado para poder realizar diversas tareas en las que se encuentra, la creación de un cuadro de mandos, para manipular los datos de manera más fácil y visual. También hemos creado una serie de gráficas para poder sacar conclusiones sobre qué país tiene más uso de las bicicletas.

### **Visual Studio Code**

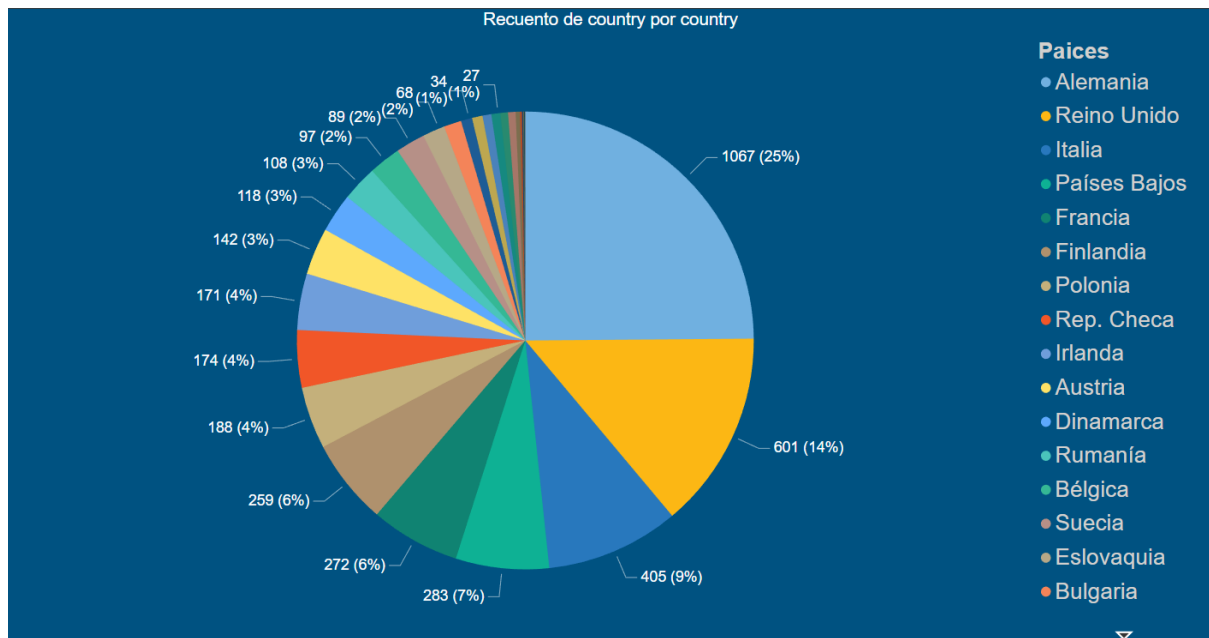
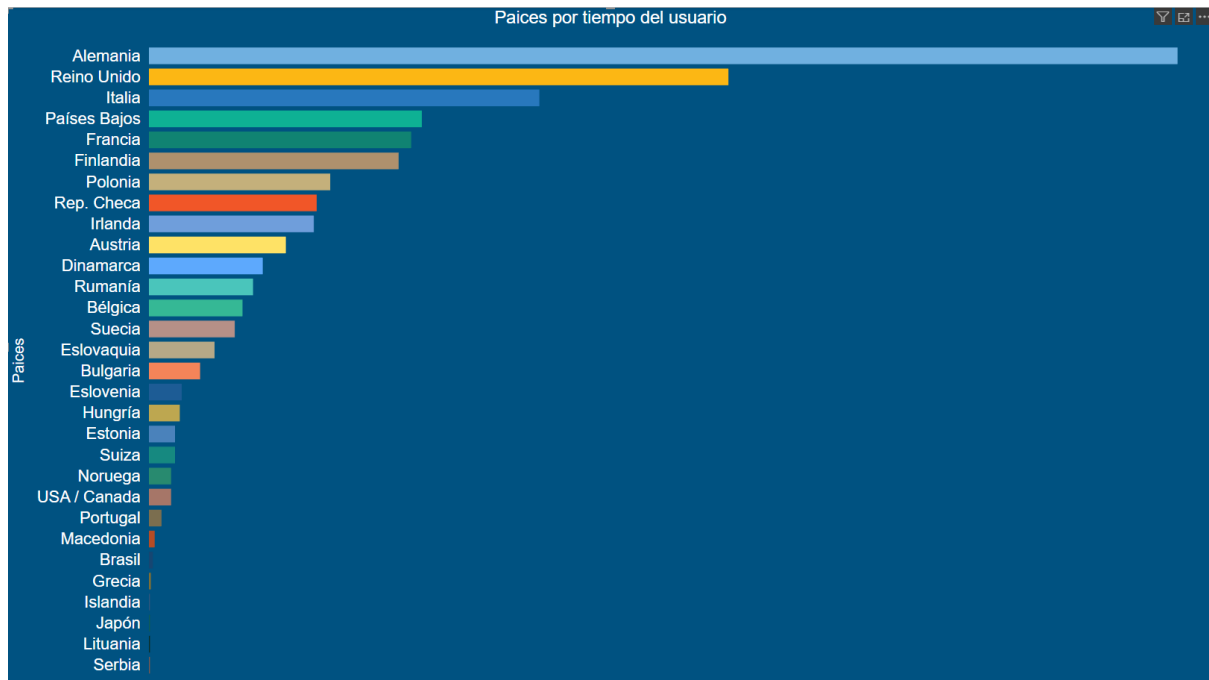
Visual Studio Code (VS Code) es un editor de código fuente desarrollado por Microsoft. Se ha convertido en una herramienta popular entre los desarrolladores debido a su enfoque en la simplicidad, su amplia gama de extensiones y su soporte para múltiples lenguajes de programación.

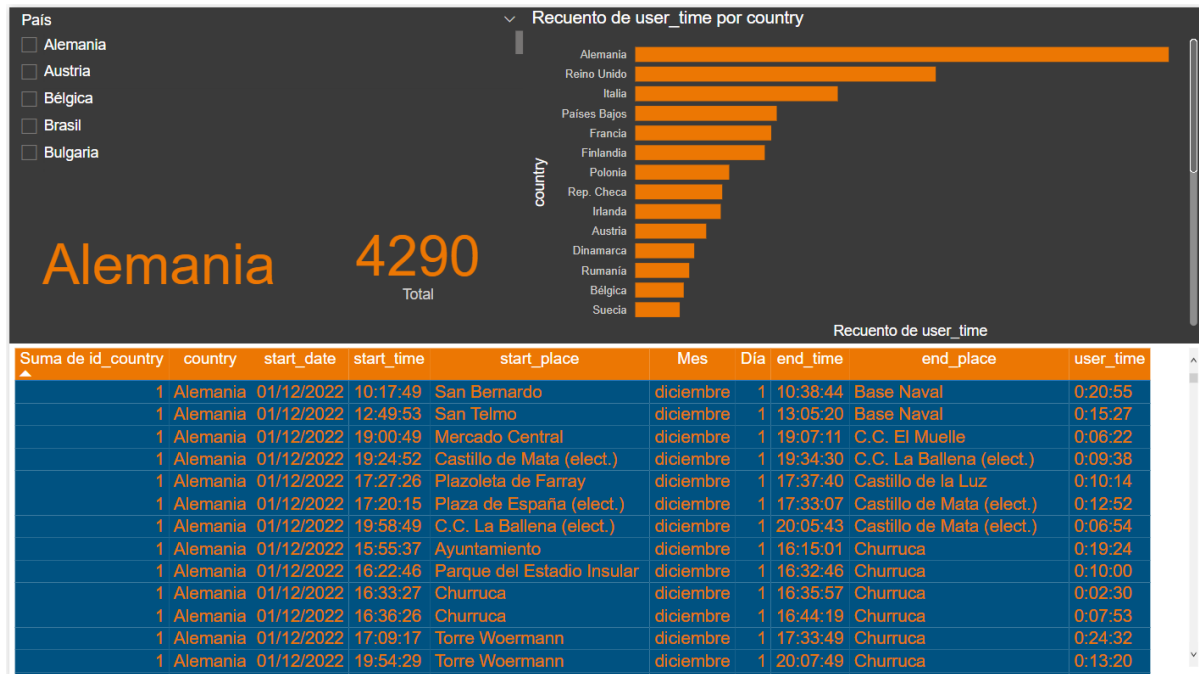
Es la herramienta que vamos a utilizar para realizar los entrenamientos, predicciones y muestreo de datos finales. También elaboramos una api para colocar en producción en la cual introducirán una serie de datos para que el modelo pueda ser iterativo.

### **Python**

Python es un lenguaje de programación de alto nivel, interpretado y de propósito general. Fue creado por Guido van Rossum y lanzado por primera vez en 1991. Python se destaca por su sintaxis clara y legible, lo que facilita la escritura y comprensión del código.

## 2. Sumario





### 3. Contexto y justificación del Trabajo

Se describe el contexto y la justificación del trabajo.

#### 3.1 Objetivos del Trabajo

Se describen los objetivos del trabajo.

#### 3.2 Enfoque y metodología

Se describe el enfoque y la metodología utilizada en el trabajo.

#### 3.3 Planificación del Trabajo



### 3.4 Recursos de sistemas utilizados

#### **Power Bi**

Power BI es una herramienta de visualización y análisis de datos desarrollada por Microsoft. Permite a los usuarios conectarse a diversas fuentes de datos, combinar información de diferentes orígenes y crear informes interactivos y paneles de control personalizados.

Con Power BI, puedes importar datos desde una amplia variedad de fuentes, como bases de datos, archivos de Excel, servicios en la nube, entre otros. Luego, puedes modelar y transformar los datos según tus necesidades utilizando el lenguaje de fórmulas DAX (Data Analysis Expressions). También puedes realizar cálculos, crear relaciones entre tablas y definir medidas para agregar y analizar los datos.

La cual hemos utilizado para poder realizar diversas tareas en las que se encuentra, la creación de un cuadro de mandos, para manipular los datos de manera más fácil y visual. También hemos creado una serie de gráficas para poder sacar conclusiones sobre qué país tiene más uso de las bicicletas.

#### **Visual Studio Code**

Visual Studio Code (VS Code) es un editor de código fuente desarrollado por Microsoft. Se ha convertido en una herramienta popular entre los desarrolladores debido a su enfoque en la simplicidad, su amplia gama de extensiones y su soporte para múltiples lenguajes de programación.

Es la herramienta que vamos a utilizar para realizar los entrenamientos, predicciones y muestreo de datos finales. También elaboramos una api para colocar en producción en la cual introducirán una serie de datos para que el modelo pueda ser iterativo.

#### **Python**

Python es un lenguaje de programación de alto nivel, interpretado y de propósito general. Fue creado por Guido van Rossum y lanzado por primera vez en 1991. Python se destaca por su sintaxis clara y legible, lo que facilita la escritura y comprensión del código.

## 4. Creación del set de datos

### 4.1 Estudio de datos necesarios

En este caso para realizar un estudio previo de los datos y manipularlos hemos utilizado la herramienta de Power BI

### 4.2 Fuente de datos

Datos libres de sitycleta(Sagulpa): <https://sagulpa.com/datos-abiertos>

En el cual hemos cogido los datos de Diciembre de 2022, ya que tiene días de fiesta y es en un tiempo no muy lejos de la actualidad ni tampoco unos años atípicos como lo fueron 2020 o 2021, marcado por la pandemia.

Aemet: [https://www.aemet.es/es/datos\\_abiertos/AEMET\\_OpenData](https://www.aemet.es/es/datos_abiertos/AEMET_OpenData)

Decidí coger los datos de tiempo de ese mismo año y mes, ya que al ser una actividad de ocio queso realiza en la calle tiene mucho factor el tiempo que pueda hacer ese día, por lo que es muy interesante ver que resultado o patrones podemos encontrar.

Dataset de tiempo

	prec	presMax	presMin	racha	tmax	tmed	tmin	velmedia	moth	day
0	3.1	1015.7	1012.0	11.4	22.3	20.6	19.0	2.8	12	1
1	1.1	1012.9	1008.8	7.8	23.0	20.7	18.4	3.3	12	2
2	7.1	1009.7	1006.5	6.7	20.1	18.9	17.7	1.7	12	3
3	0.0	1007.7	1004.9	3.9	23.0	20.1	17.2	1.1	12	4
4	0.4	1015.0	1006.1	6.9	24.2	22.0	19.9	1.4	12	5
5	0.0	1015.1	1010.9	3.6	24.9	21.7	18.5	0.3	12	6
6	0.8	1010.9	1005.6	3.3	22.7	20.8	19.0	1.1	12	7
7	0.0	1010.0	1005.7	4.7	23.7	22.0	20.2	0.6	12	8
8	0.0	1015.2	1009.2	6.1	25.0	22.4	19.9	0.8	12	9
9	0.0	1017.3	1014.1	5.0	24.8	22.4	20.0	0.8	12	10
10	0.0	1017.2	1013.9	5.0	24.6	22.6	20.6	0.6	12	11
11	0.0	1017.9	1015.3	6.1	24.1	22.0	20.0	0.8	12	12
12	0.0	1017.2	1015.0	7.2	25.0	22.2	19.4	1.1	12	13
13	0.0	1016.4	1013.7	9.2	24.6	22.2	19.8	1.7	12	14
14	4.0	1016.9	1014.0	7.8	23.2	21.6	19.9	2.8	12	15
15	0.0	1016.9	1014.0	6.9	22.5	19.9	17.3	1.9	12	16
16	0.0	1018.4	1015.5	4.4	22.9	20.8	18.7	1.1	12	17
17	0.0	1020.5	1017.1	5.8	23.0	20.6	18.3	1.4	12	18
18	0.0	1021.2	1018.7	4.7	23.3	20.6	17.8	1.1	12	19
19	0.0	1024.1	1019.6	7.2	23.7	21.0	18.2	1.4	12	20
20	0.0	1026.0	1022.6	5.8	23.5	20.6	17.7	0.8	12	21
21	0.0	1025.5	1022.2	4.7	23.8	21.0	18.3	1.1	12	22
22	0.0	1023.6	1020.7	7.5	23.5	20.8	18.0	1.4	12	23
23	0.0	1021.5	1018.5	5.8	23.1	20.2	17.4	1.4	12	24
24	0.0	1019.5	1013.1	7.5	24.7	20.8	17.0	1.7	12	25
25	0.0	1013.2	1009.9	8.3	23.2	20.4	17.7	2.5	12	26
26	0.0	1019.7	1012.0	16.9	22.5	21.1	19.7	4.7	12	27
27	0.0	1020.6	1018.1	7.8	23.2	21.8	20.5	1.1	12	28
28	0.0	1022.8	1019.3	7.5	23.7	21.4	19.1	1.4	12	29
29	0.0	1022.1	1019.5	11.4	24.9	22.5	20.1	1.9	12	30
30	0.0	1021.1	1017.9	9.7	23.8	21.4	18.9	2.5	12	31



dataset de sitycleta

	id_country	country	start_place	start_day	start_month	start_h	start_min	start_seg	end_day	end_month	end_h	end_min	end_seg	user_h	user_min	user_seg
0	18	Polonia	Plazoleta de Farray	13	12	20	59	23	14	12	8	6	9	11	06	46
1	20	Hungría	C.C. El Muelle	2	12	12	51	2	2	12	21	54	10	09	03	08
2	14	Finlandia	Plaza de España (elect.)	27	12	11	12	16	27	12	19	50	51	08	38	35
3	11	Reino Unido	Teatro Pérez Galdós	27	12	22	18	4	28	12	6	37	52	08	19	48
4	11	Reino Unido	Churruga	30	12	22	29	17	31	12	6	47	42	08	18	25
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
4279	18	Polonia	San Telmo	28	12	14	37	34	28	12	14	38	28	00	00	54
4280	15	Italia	Plaza de la feria	13	12	11	46	49	13	12	11	47	38	00	00	49
4281	11	Reino Unido	Plazoleta de Farray	4	12	21	8	3	4	12	21	8	49	00	00	46
4282	14	Finlandia	Plaza de la feria	1	12	9	11	42	1	12	9	12	27	00	00	45
4283	1	Alemania	Castillo de Mata (elect.)	21	12	21	20	44	21	12	21	21	16	00	00	32

dataset final

df_sagulpa																					
✓ 0.0s																					
	id_country	country	start_place	start_day	start_month	start_h	start_min	start_seg	end_day	end_month	...	user_h	user_min	user_seg	prec	presMax	presMin	tmax	tmed	tmin	velmedia
0	18	Polonia	Plazoleta de Farray	13	12	20	59	23	14	12	...	11	06	46	0.0	1017.2	1015.0	25.0	22.2	19.4	1.1
1	20	Hungría	C.C. El Muelle	2	12	12	51	2	2	12	...	09	03	08	1.1	1012.9	1008.8	23.0	20.7	18.4	3.3
2	14	Finlandia	Plaza de España (elect.)	27	12	11	12	16	27	12	...	08	38	35	0.0	1019.7	1012.0	22.5	21.1	19.7	4.7
3	11	Reino Unido	Teatro Pérez Galdós	27	12	22	18	4	28	12	...	08	19	48	0.0	1019.7	1012.0	22.5	21.1	19.7	4.7
4	11	Reino Unido	Churruga	30	12	22	29	17	31	12	...	08	18	25	0.0	1022.1	1019.5	24.9	22.5	20.1	1.9
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
4279	18	Polonia	San Telmo	28	12	14	37	34	28	12	...	00	00	54	0.0	1020.6	1018.1	23.2	21.8	20.5	1.1
4280	15	Italia	Plaza de la feria	13	12	11	46	49	13	12	...	00	00	49	0.0	1017.2	1015.0	25.0	22.2	19.4	1.1
4281	11	Reino Unido	Plazoleta de Farray	4	12	21	8	3	4	12	...	00	00	46	0.0	1007.7	1004.9	23.0	20.1	17.2	1.1
4282	14	Finlandia	Plaza de la feria	1	12	9	11	42	1	12	...	00	00	45	3.1	1015.7	1012.0	22.3	20.6	19.0	2.8
4283	1	Alemania	Castillo de Mata (elect.)	21	12	21	20	44	21	12	...	00	00	32	0.0	1026.0	1022.6	23.5	20.6	17.7	0.8
4284 rows x 24 columns																					

## 4.3 Recolección de los datos

Se describe la recolección de los datos.

## 5. Optimización, normalización y calidad del set de datos

Se describe la optimización, normalización y calidad del set de datos.

## 6. Desarrollo del modelo predictivo

### 6.1 Justificación del modelo seleccionado

Se describe la justificación del modelo seleccionado.

### 6.2 Descripción del modelo

Se describe el modelo utilizado.

## 7. Entrenamiento y evaluación del modelo predictivo

Se describe el entrenamiento y evaluación del modelo predictivo.





## 8. Puesta en producción (API y aplicación cliente)

Se describe la puesta en producción del modelo, incluyendo la API y la aplicación cliente.

## 9. Conclusiones y mejoras a realizar, modelos alternativos

Se describen las conclusiones y mejoras a realizar, incluyendo posibles modelos alternativos.

## 10. Bibliografía / Webgrafía

Se incluye una lista de referencias bibliográficas o de sitios web utilizados.

## 11. Presentación final

Se describe la presentación final del trabajo.