

GRUPO PELEGRINO

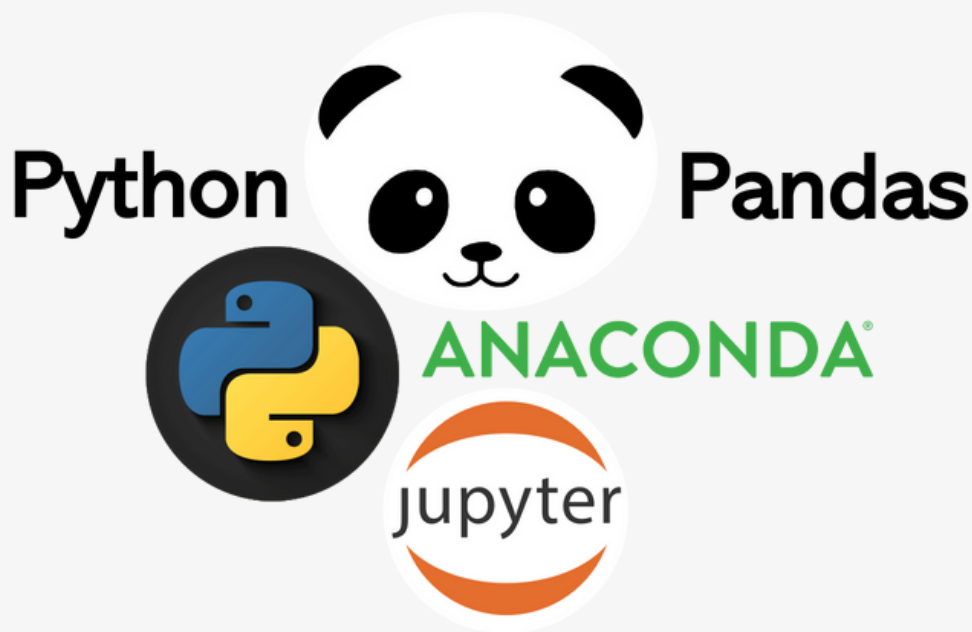
INFORME FINAL DEL PROYECTO



DESARROLLADO Y PRESENTADO POR

DILHAN BERNARD / VICTOR RENOULT / THOMAS MAAROUFI
ESTUDIANTES

CONTEXTO DEL PROYECTO



ALGUNAS INFORMACIONES SOBRE LO QUE HAY QUE HACER

Durante el curso de programación para big data, necesitamos desarrollar una aplicación que tenga como objetivo proporcionar análisis a través del procesamiento de datos.

Como parte del trabajo práctico de la asignatura tuvimos que abordar un estudio sobre un tema de interés social en el que las líneas argumentales se apoyan en tareas de tratamiento de datos. Los programas que debíamos ejecutar estaban codificados en Python, utilizando bibliotecas y técnicas que habíamos estudiado en clase.

La elección del tema de estudio nos obligó a relacionarlo con los Objetivos de Desarrollo Sostenible de las Naciones Unidas, como la igualdad de género, la energía limpia y asequible o la acción climática. Dada la naturaleza del tema, también tuvimos que centrarnos en los aspectos informáticos y de programación.

Nuestro proyecto se divide en varias tareas, como la recogida de datos, el filtrado, el preprocesamiento, la combinación, el análisis y la generación de resultados para su presentación.

El estudio debía combinar datos de al menos tres fuentes diferentes y como contrapartida debíamos tener un volumen medio, es decir, cientos o miles de registros. Los datos que teníamos que procesar debían proceder principalmente de fuentes españolas, como administraciones y organismos públicos, empresas y organizaciones no gubernamentales. Nuestros datos están disponibles en acceso libre.

DESCRIPCIÓN DEL PROYECTO

Para este proyecto hemos elegido estudiar los diferentes cambios realizados en la ciudad de Madrid y en el Ayuntamiento de Madrid. Para ello utilizamos 3 temas.

- El parque móvil del ayuntamiento de madrid
- El inventario de instalaciones fotovoltaicas en madrid
- Consumo y generación de energía del ayuntamiento de madrid

Los diferentes recursos de datos :

El parque móvil del ayuntamiento de madrid : [Aqui](#)

Consumo y generacion de energia de madrid : [Aqui](#)

El inentario de instalaciones fotovoltaicas - [Aqui](#)

NUESTRAS METAS Y OBJETIVOS

El objetivo es crear una aplicación con una buena visual, que permita hacer análisis y deducir conclusiones o tendencias sobre Madrid. En particular, sobre las medidas puestas en marcha para una migración más ecológica.

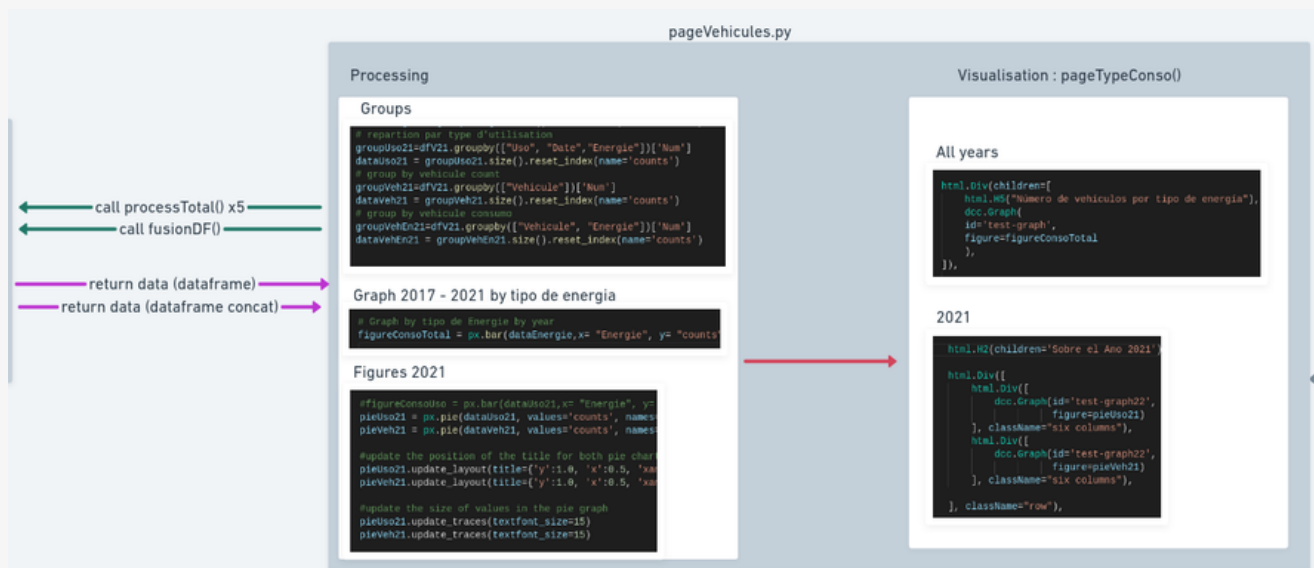


EL PLAN

- 1) Cada parte estará compuesta por 3 subpartes
 - presentación
 - desarrollo
 - análisis


```
def fusionDF(df1, df2, df3, df4, df5):
    datas = pd.concat([df1, df2, df3, df4, df5], ignore_index=True)
    return datas
```

Como hemos dicho, la función merge es una concatenación de varios dataframes (nuestros años) para crear uno global.



Para realizar nuestro análisis, una vez que nuestros marcos de datos están limpios. Hicimos grupos en función de lo que queríamos comparar y analizar

```
dataframeTotal = fusionDF(dfV17, dfV18, dfV19, dfV20, dfV21)
#print(dataframeTotal)

groupEnergie=dataframeTotal.groupby(["Energie", "Date"])['Num']
dataEnergie = groupEnergie.size().reset_index(name='counts')
# repartition par type d'utilisation
groupUso21=dfV21.groupby(["Uso", "Date", "Energie"])['Num']
dataUso21 = groupUso21.size().reset_index(name='counts')
# group by vehicule count
groupVeh21=dfV21.groupby(["Vehicule"])['Num']
dataVeh21 = groupVeh21.size().reset_index(name='counts')
# group by vehicule consumo
groupVehEn21=dfV21.groupby(["Vehicule", "Energie"])['Num']
dataVehEn21 = groupVehEn21.size().reset_index(name='counts')
```

Aquí podemos ver nuestros diferentes tratamientos para hacer una separación y preparar nuestros datos para los gráficos

```
Graph by tipo de Energie by year
figureConsoTotal = px.bar(dataEnergie,x= "Energie", y= "counts", color="Date", barmode="group")

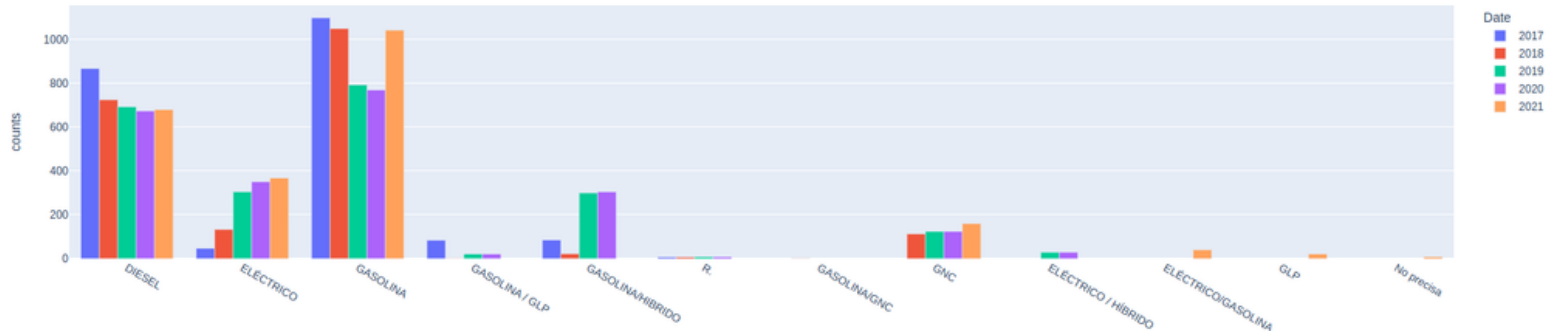
figureEnergieByVeh21 = px.bar(dataVehEn21,x= "Vehicule", y= "counts",color="Energie", barmode="group")
figureEnergieByVehTotal = px.bar(dataEnergie,x= "Energie", y= "counts", color="Energie", barmode="group")
```

Los gráficos se parametrizan con nuestros datos y se introducen en la función "page" con la estructura html de dash

ANÁLISIS

Hemos realizado un análisis global de la evolución de la energía consumida por los vehículos del parque móvil a lo largo de los años.

Número de vehículos por tipo de energía

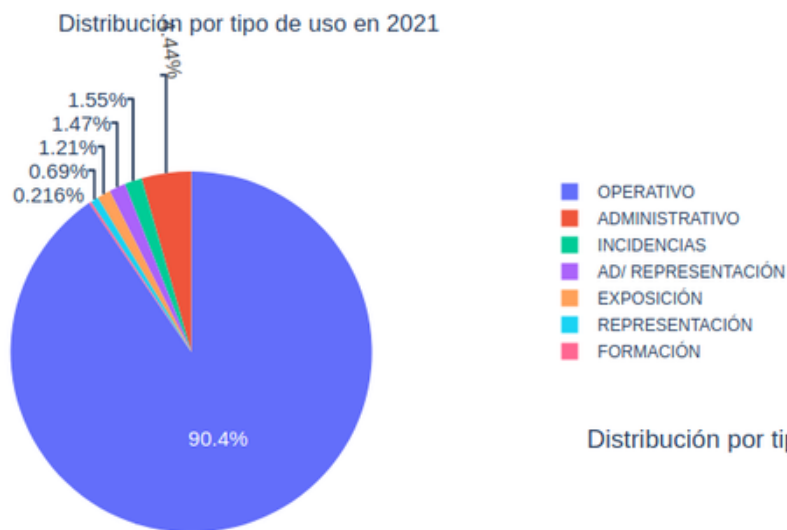


En primer lugar, se observa un descenso en el número de vehículos que consumen gasóleo de 868 (2017) a 680 (2021).

También una impresionante subida de los eléctricos con en 2017 solo 47 vehículos a 368 en 2021

La gasolina se sigue utilizando mucho, pero se observa una tendencia hacia los híbridos. (Creo que los datos de la gasolina para 2021 incluyen el híbrido).

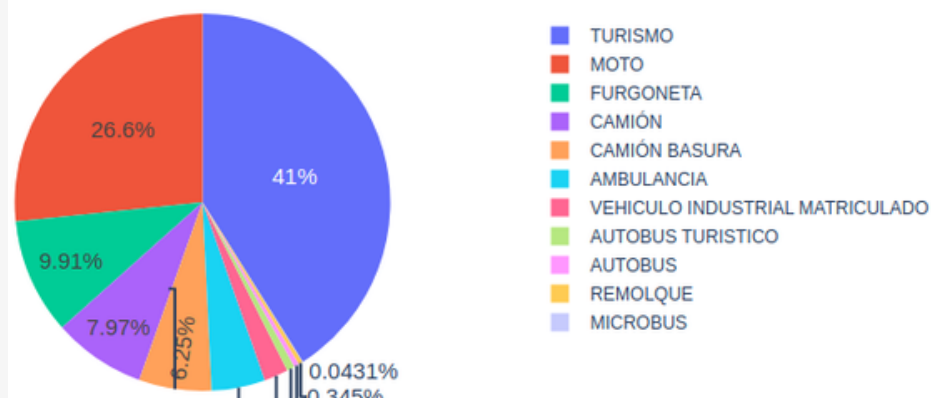
Centrémonos ahora en 2021:



Aquí tenemos nuestro diagrama sobre la distribución de vehículos por uso.

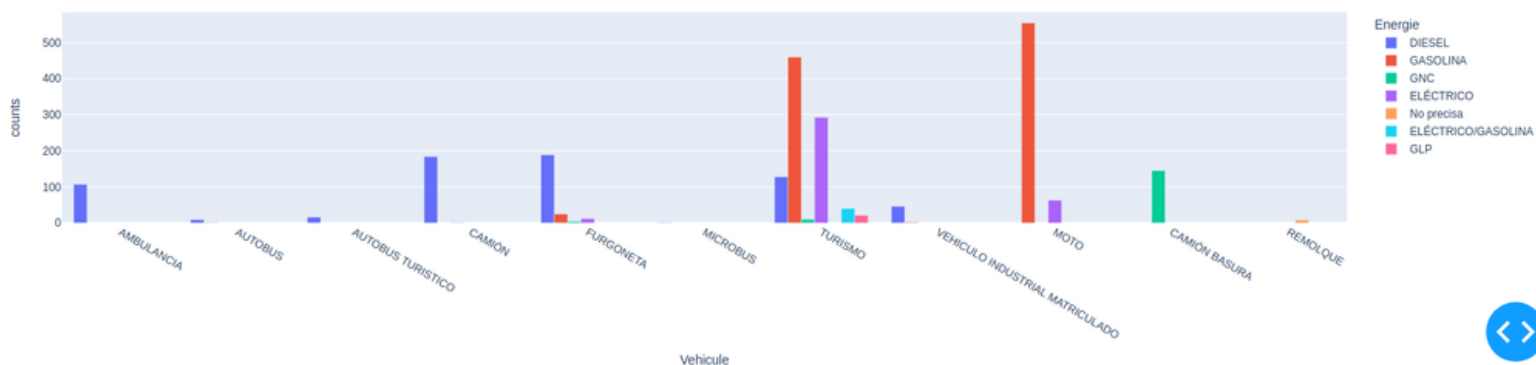
Rápidamente nos damos cuenta de que más del 90% se dedica a operaciones con 2098 vehículos, muy por delante del segundo uso, administrativo con 108 vehículos dedicados

Distribución por tipo de vehículos en 2021



Aquí tenemos nuestro diagrama de la distribución por tipo de vehículo.

Es interesante destacar que el Parque Madrid Móvil 2021 está formado principalmente por vehículos dedicados al turismo (952) y motos (618). Esto es normal dado el número de taxis, las motos son utilizadas principalmente por los guardias civiles y la policía.



Para terminar con esta parte, queríamos conocer los tipos de consumo más utilizados por las diferentes categorías de vehículos. Y podemos deducir que el diésel se sigue utilizando con bastante frecuencia en general, especialmente para las grandes categorías como ambulancias, autobuses, furgonetas, etc.

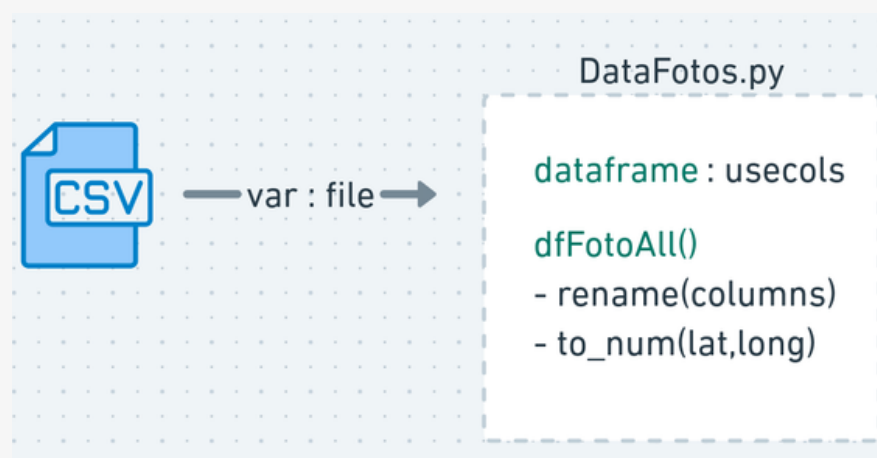
La gasolina sigue siendo muy utilizada en los vehículos de turismo y en las motos, pero hay una cierta tendencia al cambio con una buena cantidad de eléctricos e híbridos.

EL INVENTARIO DE INSTALACIONES FOTOVOLTAICAS EN MADRID

PRESENTACION

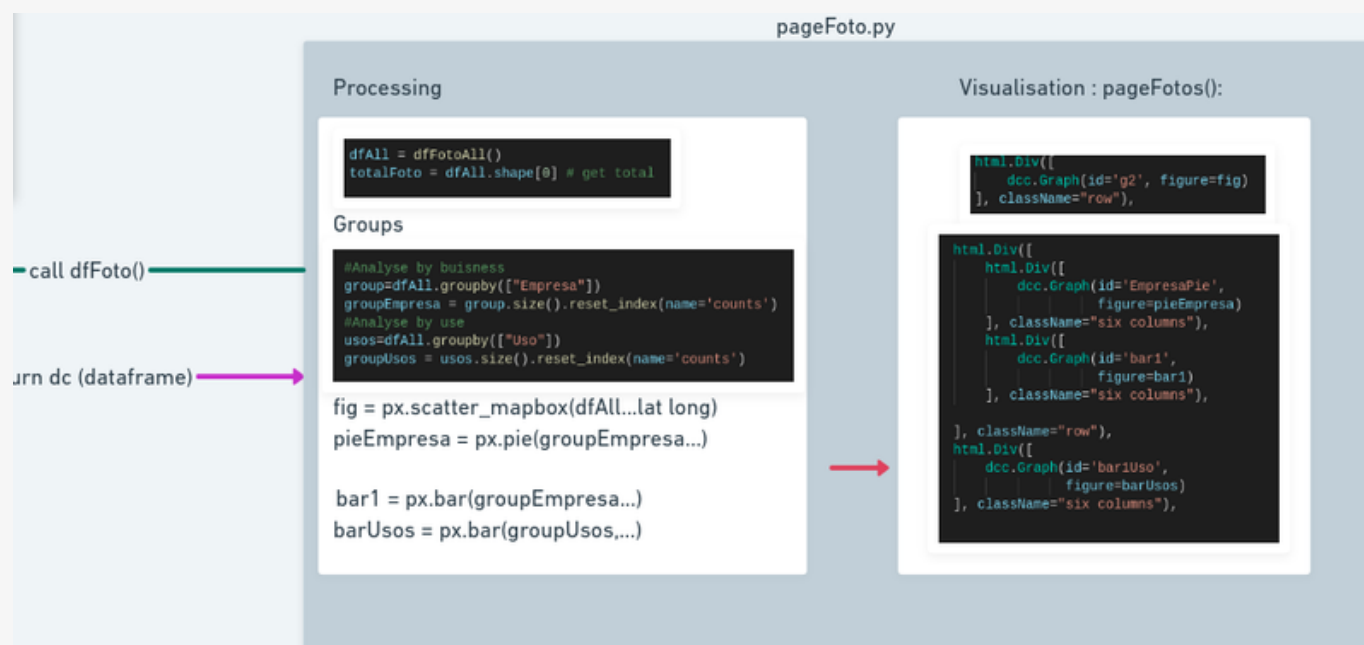
La ciudad de Madrid cuenta con varias instalaciones fotovoltaicas, en este apartado queremos analizar la distribución geográfica así como las empresas que han realizado las obras.

DESARROLLO



El sitio data.es nos ofrece un único archivo csv que hemos transformado en un marco de datos. Seleccionamos y renombramos las columnas que consideramos más útiles para un análisis interesante.

Nos dimos cuenta de que las coordenadas gps no se consideraban números reales, así que las convertimos.



Después de tener nuestro dataframe creamos varios groupBy de acuerdo a nuestras elecciones de análisis. Una vez listos, nuestros dataframes son usados en los diferentes gráficos de dash y empujados a la página web.

ANÁLISIS

Gracias a las coordenadas y a un openstreetmap integrado en dash pudimos tener una visual sobre las diferentes ubicaciones de las instalaciones.

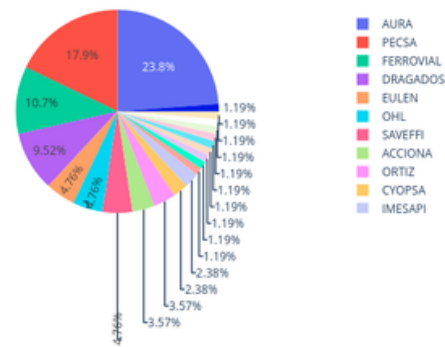
Inventory of the photovoltaic installations of Madrid



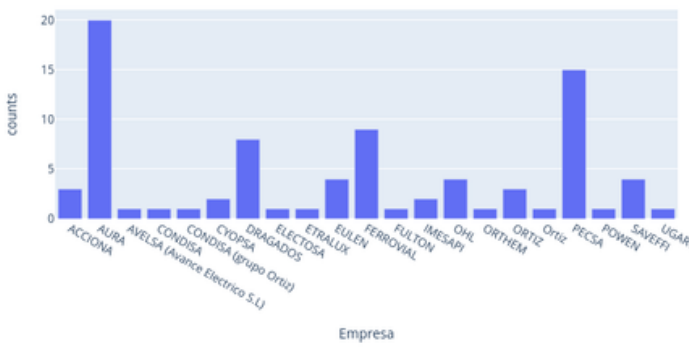
Podemos ver bastante cobertura desde Madrid.

Número de instalaciones : 105

Instalaciones por empresa (Gráfico circular)



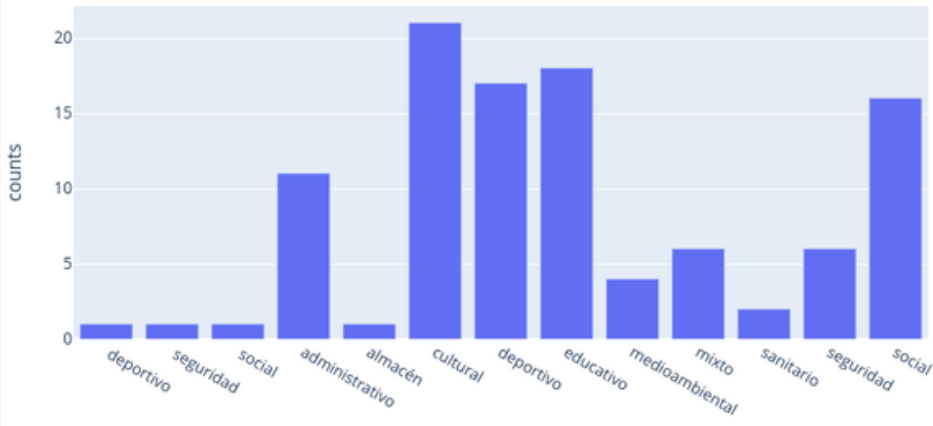
Instalaciones por empresa (Gráfico de barras)



El ayuntamiento ha convocado a diferentes empresas, pero queríamos saber cuáles eran las más convocadas o las que habían ganado más concursos. Resulta que las empresas AURA y PECSA dominan las instalaciones.

Cada instalación tiene un uso específico, aquí podemos ver el número de instalaciones por sector. También podemos deducir que el sector cultural es el que más instalaciones tiene con 21. Podemos confirmar la tendencia de la ciudad de Madrid a realizar una transición ecológica en estos sectores, cultural, deportivo, educativo y de administración social.

Instalaciones por uso (Gráfico de barras)

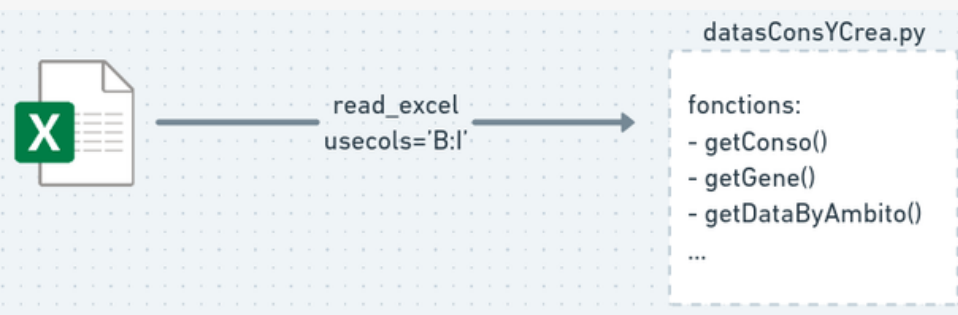


CONSUMO Y GENERACIÓN DE ENERGÍA DEL AYUNTAMIENTO DE MADRID

PRESENTACION

El consumo y la generación de energía es un dato importante, a la vez que crucial e interesante. En este estudio queremos analizar la evolución de la generación y producción de energía en la ciudad de Madrid.

DESARROLLO



Aquí no tenemos un CSV sino un XLSX, el tratamiento sigue siendo el mismo, obtendremos las columnas que nos interesan. Omitiremos las filas y columnas que estén vacías, ya que se trata de una tabla excel.

El cambio es que nuestro archivo agrupa directamente varios años diferentes, es una especie de concat.

Obtener sólo una parte de los datos para hacer nuestro análisis, hemos creado bucles que, dependiendo del año, lanzarán una función específica que recibirá lo que necesitamos.

```
def getYearTs(year1, year2): #Récup toute les dates et les mets dans une list (en str)
    year = []
    for yearR in range(year1, year2 + 1):
        yS = yearR
        sYear = str(yS)
        year.append(sYear)
        yearR = yearR+1
    return year
# Permet de récup les valeurs pour le premier graph
#-----\-/-----
def getConsoT(year1, year2):
    conso = []
    for year in range(year1, year2 + 1):
        c = getConso(year)
        conso.append(c)
        year = year+1
    return conso

def getGeneT(year1, year2):
    gene = []
    for year in range(year1, year2 + 1):
        g = getGenere(year)
        gene.append(g)
        year = year+1
    return gene
```

```
def getConso(year):
    conso = df1t.loc[(df1t['AÑO'] == year) & (df1t['CLASE'] == "Consumida"), 'CANTIDAD'].sum()
    return conso
def getGenere(year):
    geno = Gene2020 = df1t.loc[(df1t['AÑO'] == year) & (df1t['CLASE'] == "Generada"), 'CANTIDAD'].sum()
    return geno
```

Estas funciones nos permiten recuperar ordenando selectivamente el marco de datos global.

ANÁLISIS

El objetivo es analizar el consumo y la producción de energía a lo largo de 3 años para el conjunto del municipio de Madrid con el fin de interpretar los cambios, destacar la "zona" que más consume y produce con el tipo de producción utilizado. Gracias a este análisis, podremos destacar las diferentes tendencias de cambio (hacia una producción más ecológica, por ejemplo).

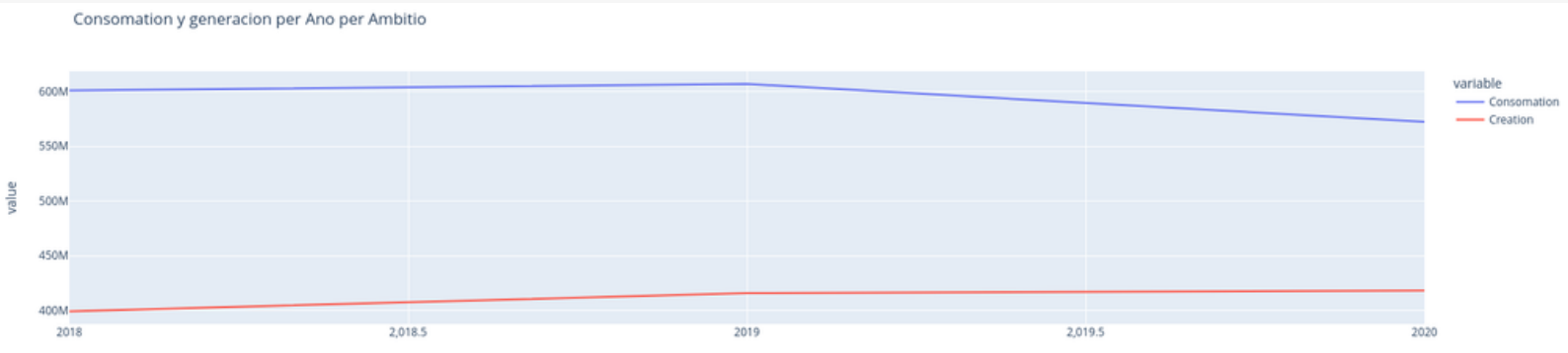


Table resumen en KWH

Year	Consomation	Creation	difference
2018	600842034.3272165	399347302	201494732.3272165
2019	606835400.7966	415954022.96999997	190881377.82660002
2020	572156462.2372	418361036.22569996	153795426.01150006

Gracias a este gráfico y a la tabla podemos ver claramente una disminución del consumo energético global de Madrid pero un aumento de su producción energética, lo que se confirma con una diferencia decreciente. Pero eso no es todo, para entrar en más detalle nos gustaría tener los detalles por ambito.

Consomation per Ano per Ambito

Ambito	2018	2019	2020
Parque Tecnológico de Valdemingómez	135798360.88	124221631.685	132210414.145
Edificios y Centros	305985303.097946	330459024.02	292696134.268
Parque móvil	20569645.3492705	15978990.437499998	14235055.02420003
Parques y viveros	4046087	3875968.6541	4025900.8
Alumbrado e instalaciones	134442638	132299786	128988958

Generacion per Ano per Ambito

Ambito	2018	2019	2020
Parque Tecnológico de Valdemingómez	399299786	413710210	415549294
Edificios y Centros	47516	2238885.95	2773916.0757
Parque móvil	0	0	0
Parques y viveros	0	4927.02	37826.15
Alumbrado e instalaciones	0	0	0

Estas imágenes nos muestran varias cosas:

- El parque tecnológico de Valdemingomez consume mucho, pero también produce cada vez más.
- Los Edificios consumen cada vez menos pero por el contrario producen cada vez más, con una explosión entre 2018 y 2019.
- A partir de 2019 los parques y vevieros comenzaron a producir energía (principalmente gracias a la fotovoltaica).
- El parque móvil consume cada vez menos.

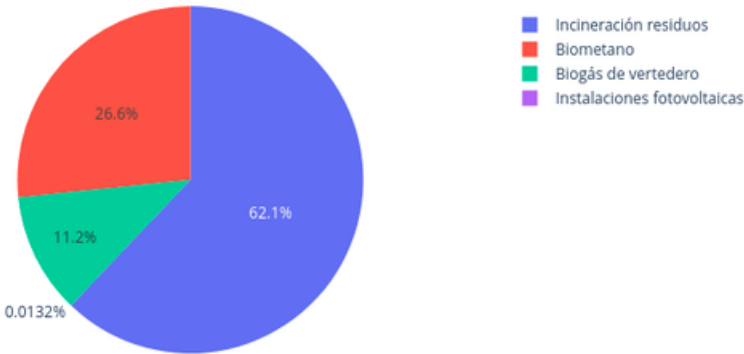
Centrémonos ahora en los tipos de energía generados por la ciudad por año desde 2018 hasta 2020 para deducir una tendencia.

De los 399347302 KWH producidos, en 2018, el 62,1% de la producción energética procede de la incineración de residuos.

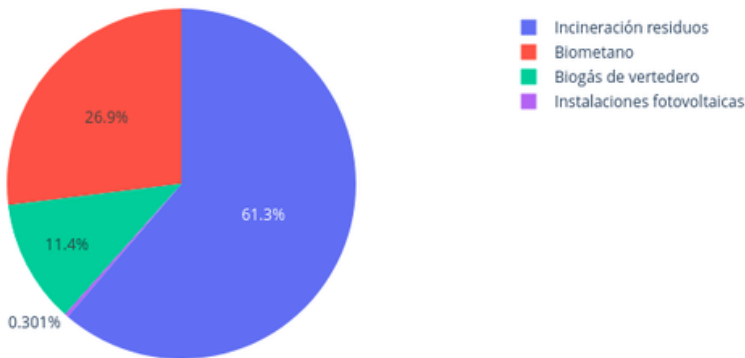
Sin embargo, es importante subrayar que el biometano y el biogás de vertedero ya representaban una buena parte de la producción "menos contaminante" de Madrid.

Obsérvese que la fotovoltaica seguía siendo mínima.

2018 generation



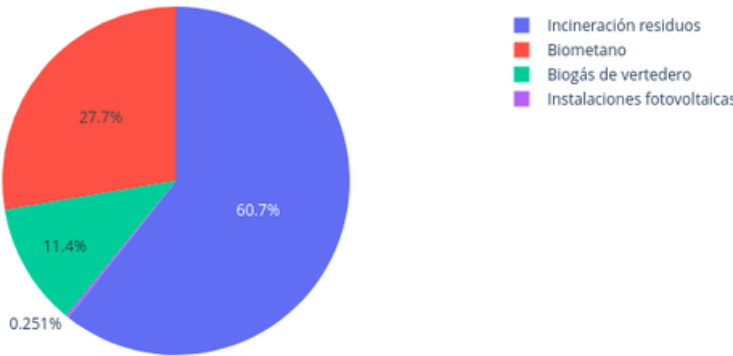
2019 generation



En 2019 sobre 415954023 KWH producidos se observa una pequeña disminución de los residuos de incineración, un ligero aumento del biometano y del biogás. Pero sobre todo una apariencia real de las fotovoltaicas.

En 2020 sobre 418361036 KWH producidos todavía observamos una pequeña disminución de los residuos de incineración, un aumento interesante de biometano. Pero extrañamente una disminución de las fotovoltaicas.

2020 generation





INFORME DE ANÁLISIS

Madrid es una de las ciudades más grandes y poderosas de Europa. Hemos asistido a una evolución que sigue una tendencia eco-responsable con la disminución de la producción de energía contaminante, la llegada de los vehículos híbridos y eléctricos. Una evolución de su parque móvil hacia un combustible más ecológico.

Lo que sí es cierto es que Madrid está poniendo los medios necesarios para integrar esta visión de energía verde y cada vez menos contaminante.

ORGANIZACIÓN DEL PROYECTO

TAREA	PRINCIPIO	FIN
INVESTIGACIÓN INVESTIGACIÓN Y TRATEAR	OCTUBRE	NOVIEMBRE
ENTRAGA 1 FASE 1	NOVEMBER	
FINAL ENTRAGA FASE 2	DICIEMBRE	

ANNEXO

DEPENDENCIAS

La aplicación requiere :

- Pandas
- dash
- dash_bootstrap_components
- openpyxl

otras dependencias normalmente ya están instaladas pero he creado un request.t si es necesario

PROYECTO

Repositorio github : https://github.com/ArK4nes/Prog_BigData_Dilhan

MINDMAP DEL APPLICATION

