

Tetuan City Power Consumption

Uri Jared Gopar Morales, ITC A01709413, Tecnológico de Monterrey Campus Querétaro.

Abstracto- En el siguiente documento se presentará el análisis de los datos sobre el consumo de energía en la ciudad de Tetuán en Marruecos por medio de un dataset.

I. INTRODUCCIÓN

Tetuán se encuentra en el norte de Marruecos a orillas del Mar Mediterráneo. Esta ciudad se caracteriza por su rica cultura, fue reconocida como Patrimonio de la Humanidad por la UNESCO. Sus principales fuentes de ingresos son la agricultura, pesca, turismo y el comercio. A lo largo del tiempo su población fue aumentando drásticamente tanto por españoles como extranjeros.

1. Evolución de la población por Distrito y Barrio											
	Población a 1 de enero										
	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
06. Tetuán	156.433	155.764	155.884	152.523	158.880	152.945	152.789	155.967	157.997	161.219	159.849
06.1. Bellas Vistas	29.090	29.362	29.074	28.526	28.074	28.270	28.359	28.750	29.146	29.895	29.465
06.2. Cuatro Caminos	34.473	34.244	34.262	33.620	33.036	33.627	33.638	34.254	34.608	35.782	34.977
06.3. Castilletes	20.002	20.343	20.397	19.960	19.639	19.799	19.962	20.363	20.519	20.805	20.541
06.4. Almorana	22.187	22.179	22.119	21.897	21.691	21.769	21.962	22.421	22.642	23.083	22.738
06.5. Valdecañales	25.221	25.286	25.494	24.921	24.583	24.932	25.274	25.571	26.021	26.545	26.625
06.6. Berenguete	24.000	24.450	24.357	23.783	23.637	24.198	24.314	24.608	25.007	25.843	25.683
Ciudad de Madrid	3.208.881	3.237.937	3.215.633	3.168.130	3.145.991	3.165.883	3.182.175	3.221.824	3.266.126	3.304.730	3.312.210

2. Números índice por Distrito y Barrio. Base 2010 = 100											
	Población a 1 de enero										
	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
06. Tetuán	100	99.6	99.5	97.5	98.4	97.5	98.3	99.7	101.0	101.1	100.2
06.1. Bellas Vistas	100	99.9	99.3	97.7	96.7	97.3	97.8	97.2	98.5	101.0	99.6
06.2. Cuatro Caminos	100	99.3	99.4	97.6	96.7	97.6	98.4	99.4	100.4	102.1	101.0
06.3. Castilletes	100	100.2	100.5	97.8	96.8	96.6	97.5	98.3	100.1	101.6	100.2
06.4. Almorana	100	100.2	99.9	98.9	97.9	98.3	99.3	101.3	102.3	103.9	102.7
06.5. Valdecañales	100	100.2	101.0	98.8	97.4	98.8	99.9	101.3	103.1	105.2	105.5
06.6. Berenguete	100	99.8	99.4	97.1	96.5	98.8	99.2	100.4	102.1	105.5	104.7
Ciudad de Madrid	100	99.0	98.3	96.8	96.1	96.8	97.3	98.5	99.9	102.0	101.3

Causando como resultado un aumento en la demanda energética, esto afecto directamente a los sectores clave como la industria, comercio y el medio ambiente.

Como precaución se registraron los datos del consumo de energía del año 2017, se midieron diferentes atributos los cuales fueron, la fecha con registros cada 10 minutos, temperatura, humedad, velocidad del viento, flujos difusos generales y flujos difusos, se dividió al territorio en 3 zonas diferentes para poder saber el consumo de energía de estas.

II. MANEJO DE DATOS

El dataset cuenta con 52,417 por cada día del año que se tomaron las medidas, para tener una vista panorámica de las condiciones en las cuales el consumo de energía es mayor en las zonas.

Sin embargo, nos enfocaremos en predecir el consumo de la primera zona y los meses de julio y agosto de 2017. Dejándonos (# de datos). Se decidió la zona 1 por su principal consumo de energía a través del año, ya que es mayor que el de las zonas 2 y 3. Al seleccionar los meses fueron los que más consumo tuvieron durante ese año, nos ayudaran a poder predecir el consumo máximo de los siguientes años, únicamente utilizaremos 2 meses para tener un mejor resultado de las predicciones mejor sin muchos sesgos por aquellos meses que el consumo de energía fue menor, pero que otras variables afecto dicho consumo.

La zona con mayor consumo total de energía es: Zone 1 con un consumo de 1695393977.06 unidades.
Los dos meses con mayor consumo en Zone 1 son:
Mes: 8, Consumo: 162646686.26 unidades
Mes: 7, Consumo: 159952055.28 unidades

Teniendo definido nuestra área de estudio, se realizo los ajustes al dataset, los cuales fueron la normalización de las variables a estudiar, para que todos los valores se encuentren en un rango de [0 a 1], facilitando de esta manera el análisis y dejando los posibles errores de que las variables más grandes sesguen a las pequeñas causando esto discrepancia en los datos, la formula que utilice para dicha actividad fue la de normalización Min-Max:

$$X_{normalized} = \frac{x - x_{min}}{x_{max} - x_{min}} = \epsilon [0,1]$$

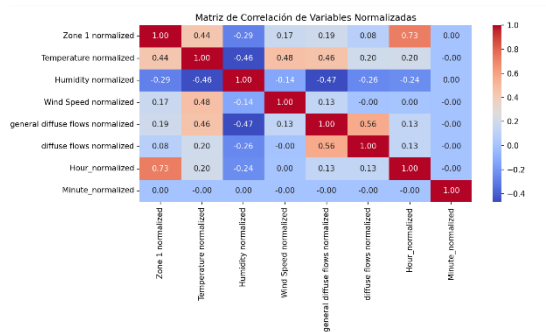
Teniendo los valores normalizados, ahora vamos a estandarizar dichos datos, esta función nos va a ayudar con nuestra

distribución de datos los cuales los va a centralizar.

$$X_{standardized} = \frac{x - \text{mean}(x)}{\text{std}(x)}$$

III. REGRSIÓN LINEAL

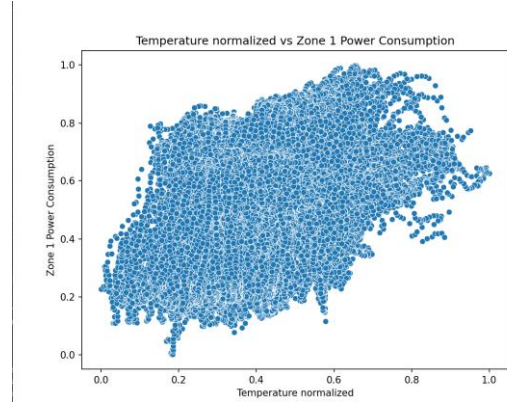
Escogí como punto de partida la regresión lineal múltiple, para entender un poco el contexto de las variables y su comportamiento gráfico mi matriz de Correlación de Variables Normalizadas. En la que se puede apreciar como es que las diferentes variables tienen un efecto en nuestra variable dependiente que sería la zona 1, por lo que si el numero de la matriz es 1 indica una correlación positiva lo que hace que si esa variable sube nuestra variable dependiente también lo hará.



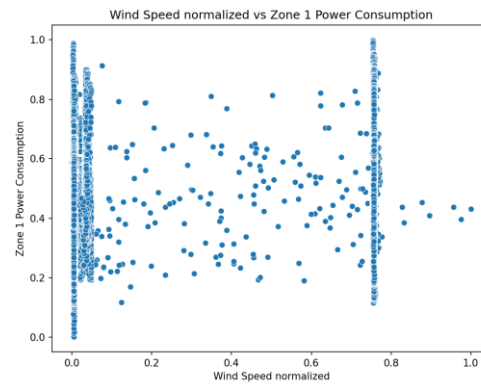
Como un resumen rápido de esta matriz, podemos observar que las variables que más influyen en el consumo de energía son: las horas, la temperatura, velocidad del viento y los flujos difusos generales. Obteniendo esto ya sabemos que variables nos ayudaran el correcto entrenamiento de nuestro modelo para una mejor predicción. Las demás variables no afectan mucho.

Vamos a ver un poco como es el comportamiento de los datos y su influencia en el consumo de energía.

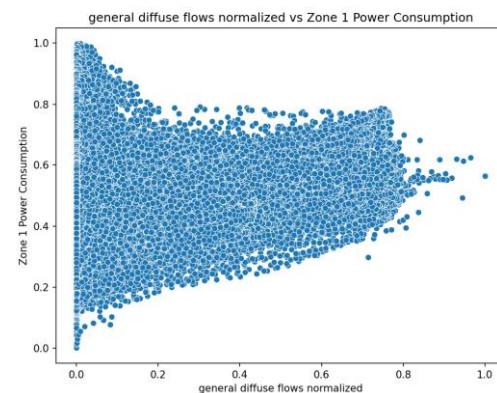
Relación entre temperatura y zona 1.



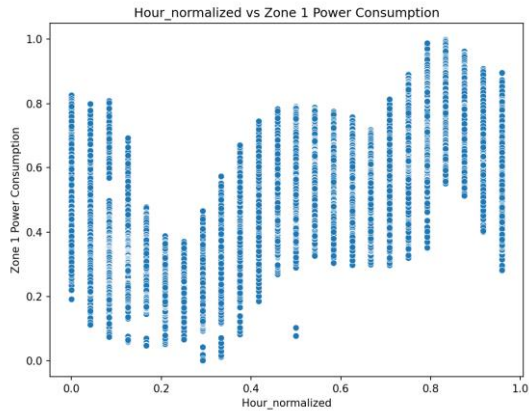
Relación entre velocidad del viento y zona 1



Relación de flujos difusos generales y zona 1



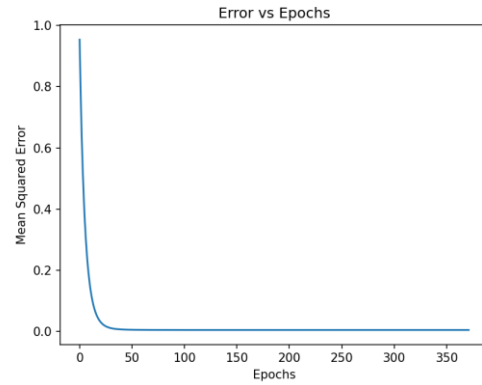
Relación hora y zona 1



Estas gráficas me ayudaron a poder comprender que el comportamiento de la dispersión no es perfectamente lineal, sin embargo, me percate como es que todas las variables tienen influencias en la variable dependiente que en este caso es el total de consumo de energía de la zona 1. Dicho esto, el modelo de regresión lineal como punto de partida me ayudará para poder predecir el consumo de energía, pero al ser este muy simple tendrá algunas limitaciones.

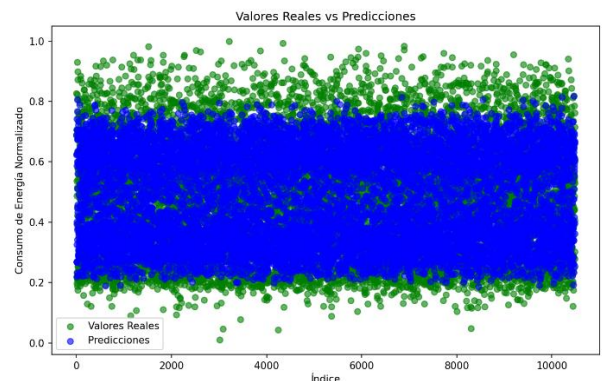
Para implementar esta regresión lineal, separe mi dataset en 80% entrenamiento y 20% valores de prueba, con la finalidad de poder ajustar los parámetros al modelo y de esta manera poder aprender las relaciones entre las variables independientes y la variable dependiente. Para implementar esto hice uso del algoritmo de gradiente descendiente, para poder ajustar los parámetros del modelo, los cuales se ajustan por la variable de aprendizaje (alfa). Y de esta manera minimizar la diferencia entre las predicciones y los datos reales.

Después de ajustar el modelo, se calcula el MAE (Mean Absolute Error), que mide la magnitud media del error absoluto entre las predicciones y los valores reales. Esto proporciona una medida de la precisión del modelo.



Después de entrenar este modelo, se utilizan los parámetros ajustados para predecir los elementos de prueba.

Por último, gracias a la función de coeficiente de determinación R^2 , nos ayuda a medir que tan buenas fueron las predicciones de nuestro modelo. Mientras más cercano a 1 sea el valor mejor predijo nuestro modelo. En este caso mi modelo tiene una R^2 de 0.69% el cuál, aunque es un poco baja tiene sentido, ya que al ser esta una regresión lineal no puede predecir los valores más altos o bajos, sin embargo, este modelo puede mejorar su predicción al pasarlo a una red neuronal.



Bibliografía

- Explotación del Padrón Municipal de Habitantes 2021. (2021). *Distrito 06 - Tetuán*. Obtenido de <https://www.madrid.es/UnidadesDescentralizadas/UDCEstadistica/Nuevaweb/Publicaciones/Padr%C3%B3n%20Municipal%20de%20Habitantes/2021/dto06.pdf>
- Salam, A. &. (Diciembre de 2018). *Comparison of Machine Learning Algorithms for the Power Consumption Prediction:-Case Study of Tetouan city*“. Obtenido de <https://archive.ics.uci.edu/dataset/849/power+consumption+of+tetouan+city>
- Universidad Piloto de Columbia. (2020). *estudiarvirtual.unipiloto.edu.co*. Obtenido de <https://estudiarvirtual.unipiloto.edu.co/blog/evaluar-modelo-de-regresion-logistico#:~:text=El%20cuadrado%20de%20R%20se,predicci%C3%B3n%20y%20el%20valor%20real.>