Aprenentatge Computacional

Práctica 1: ANÁLISIS DE DATOS Y REGRESIÓN LINEAL

Grup GPA603-1530

Mario González 1566235
Ferran Bernabé 1564845
Daniel Gutiérrez 1563389

Índice de contenidos

Explicación de la BBDD a analizar

Explicación de la BBDD

- Visitas a Yellow Stone National Park
- Comparación con otros factores:
  - Temperatura
  - Precipitación / climatologia
  - Factores economicos
  - Factores sociales

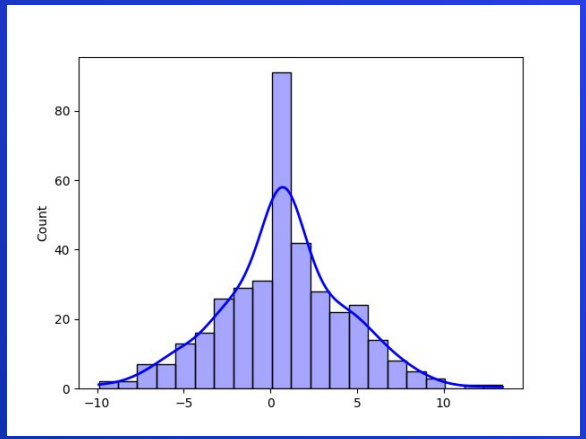
Atributos de la BBDD

Visitas	Temperatura	Climatología	Factores sociales-económicos
Recreation Visits	- LowestTemperature(F) - HighestTemperature(F) - WarmestMinimumTemperature(F) - ColdestMaximumTemperature(F) - AverageMinimumTemperature(F) - AverageMaximumTemperature(F) - MeanTemperature(F)	- TotalPrecipitation(In) - TotalSnowfall(In) - Max 24hrPrecipitation(In) - Max 24hrSnowfall(In)	- 3month Percent Change Airfare Costs - 3month Percent Change Food Away From Home Costs - 3month Percent Change Gasoline Costs - 3month Percent Change Jet Fuel Costs - Consumer Price Index - Consumer Sentiment Index - Unemployment Rate

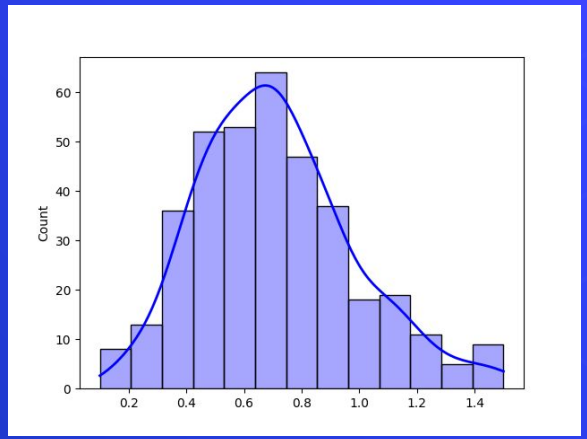
Análisis de los datos

Distribuciones Gaussianas

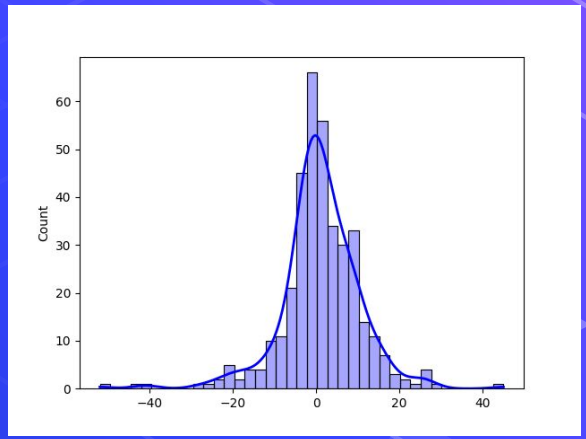
- En primer lugar hemos observado que atributos tienen una distribución Gaussiana:



3month Percent Change Airfare Costs



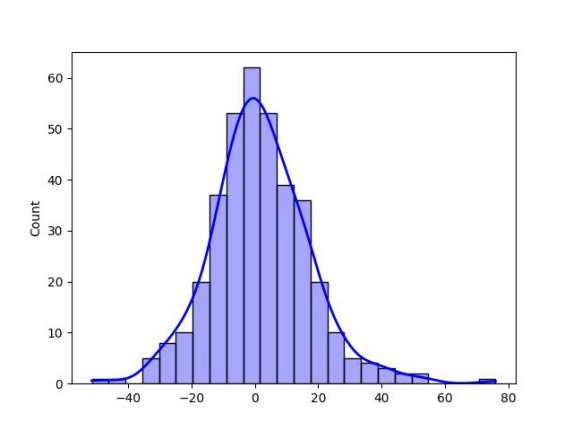
3month Percent Change Food Away From Home Costs



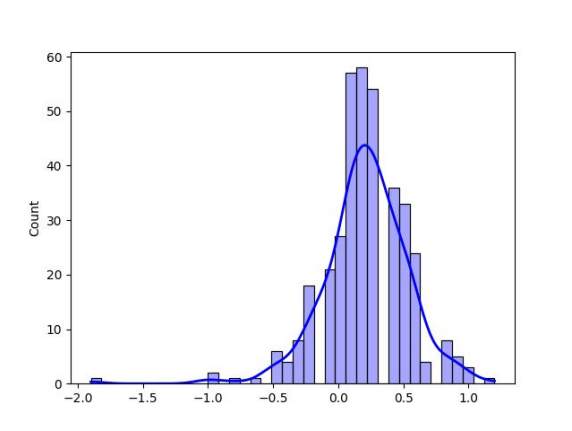
3month Percent Change Gasoline Costs

Distribuciones Gaussianas

- En primer lugar hemos observado que atributos tienen una distribución Gaussiana:



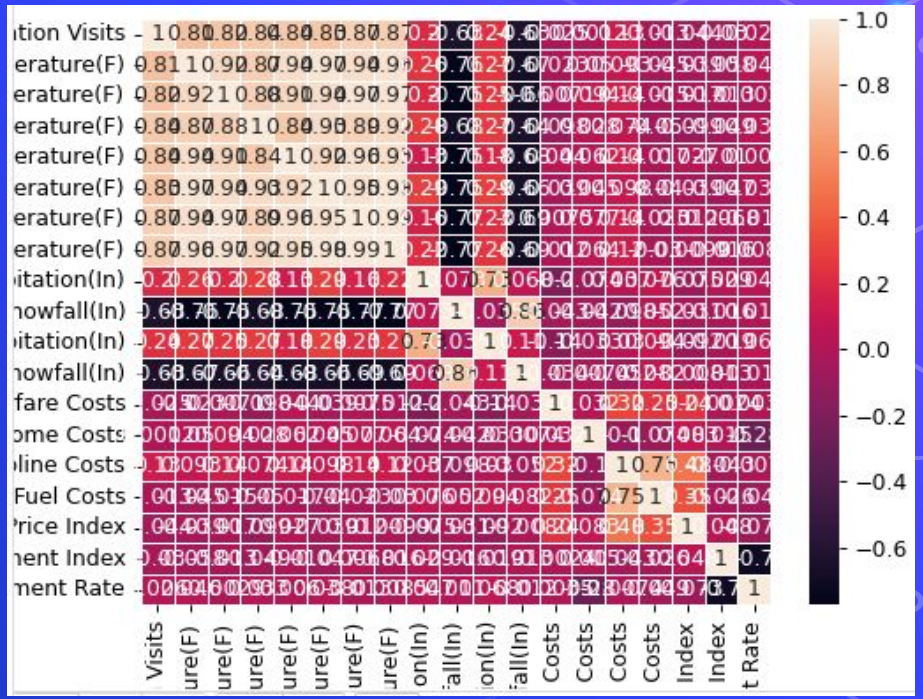
3month Percent Change Jet Fuel Costs



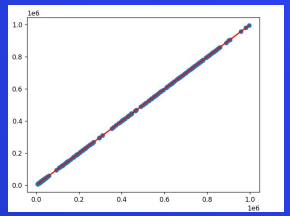
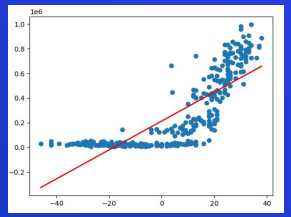
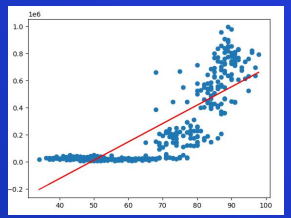
Consumer Price Index

Matriz de correlación

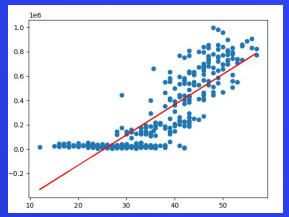
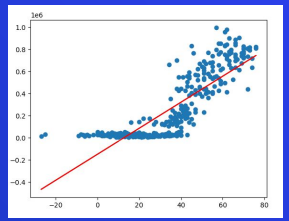
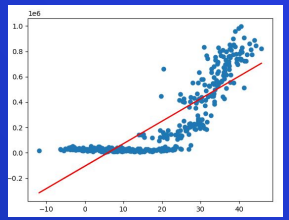
- En segundo lugar, hemos observado la matriz de correlación:
- Hipótesis: Los atributos de temperatura tienen una relación causal con las visitas.



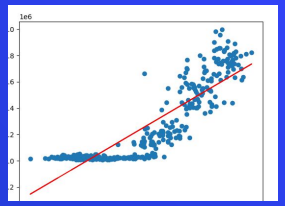
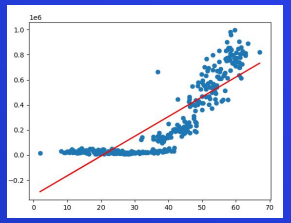
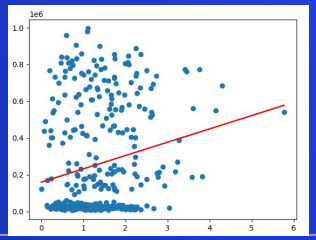
Cálculo del MSE

Atributo	MSE	MSE Normalizado
Recreation Visits 	0.000000	0.000000
LowestTemperature 	28940377249.70259	0.3409128671277541
HighestTemperature 	27753677157.985798	0.32693373592302527

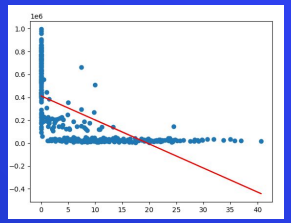
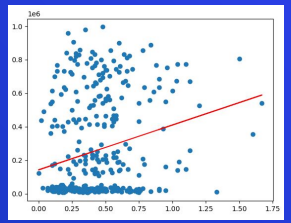
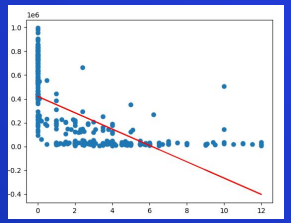
Cálculo del MSE

Atributo	MSE	MSE Normalizado
WarmestMinimumTemperature 	24623723241.10622	0.2900633954096815
ColdestMaximumTemperature 	25478142991.34341	0.30012831903768755
AverageMinimumTemperature 	26259642721.500072	0.30933425686526067

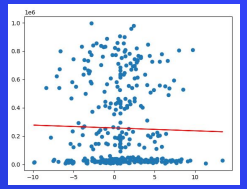
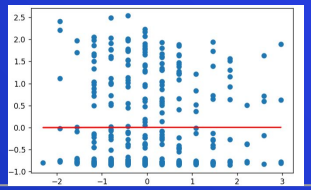
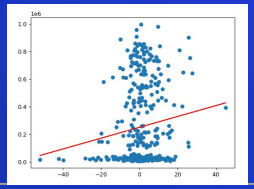
Cálculo del MSE

Atributo	MSE	MSE Normalizado
AverageMaximumTemperature 	20030670686.00058	0.2359579944358358
MeanTemperature 	21234794994.757767	0.2501423800412746
TotalPrecipitation 	81586862197.06857	0.9610797700242644

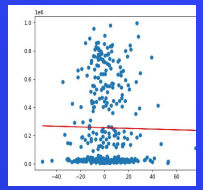
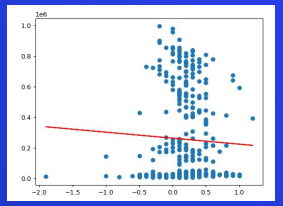
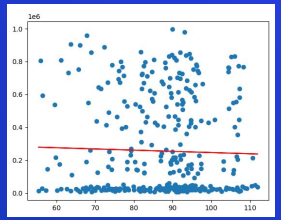
Cálculo del MSE

Atributo	MSE	MSE Normalizado
TotalSnowfall 	51063852220.06794	0.6015237505969343
Max 24hrPrecipitation 	80133020916.02104	0.9439537597033142
Max 24hrSnowfall 	51321432040.92007	0.6045579983706689

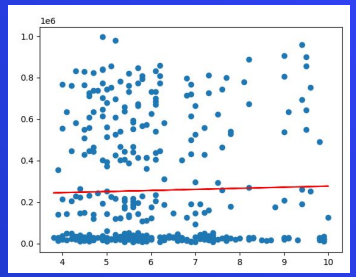
Cálculo del MSE

Atributo	MSE	MSE Normalizado
3month Percent Change Airfare Costs 	84838850298.31592	0.9993876530866032
3month Percent Change Food Away From Home Costs 	84890706659.5314	0.9999985124623553
3month Percent Change Gasoline Costs 	83471220190.70598	0.9832771961588057

Cálculo del MSE

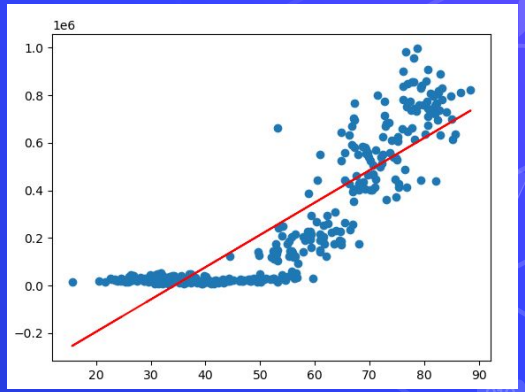
Atributo	MSE	MSE Normalizado
3month Percent Change Jet Fuel Costs 	84875601572.70674	0.9998205770327933
Consumer Price Index 	84727417784.11115	0.9980749964622257
Consumer Sentiment Index 	84814425923.95772	0.9990999379881297

Cálculo del MSE

Atributo	MSE	MSE Normalizado
Unemployment Rate 	84832147714.52428	0.9993086977558603

Atributos escogidos

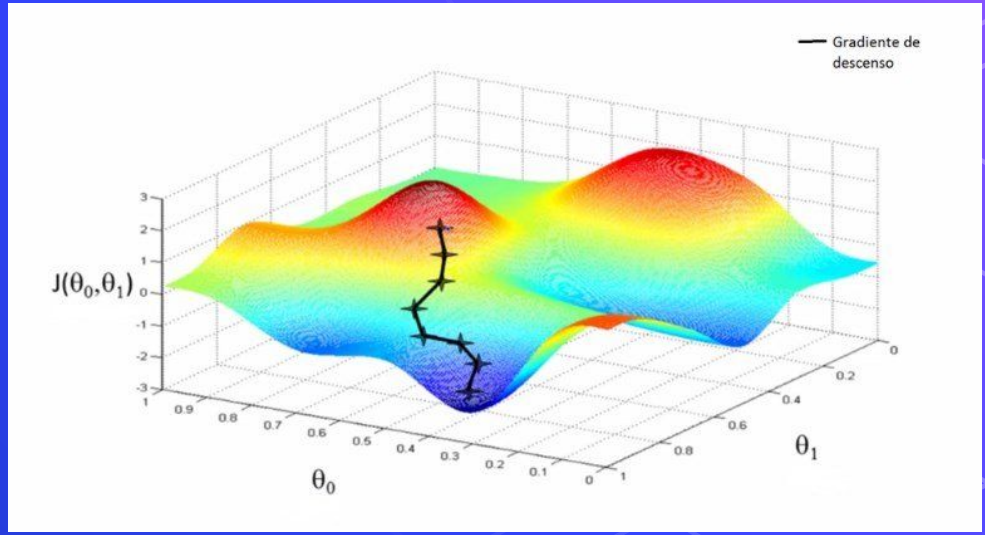
- LowestTemperature
- HighestTemperature
- WarmestMinimumTemperature
- ColdestMaximumTemperature
- AverageMinimumTemperature
- **AverageMaximumTemperature**
- **MeanTemperature**



El descenso de gradiente

Descenso de gradiente

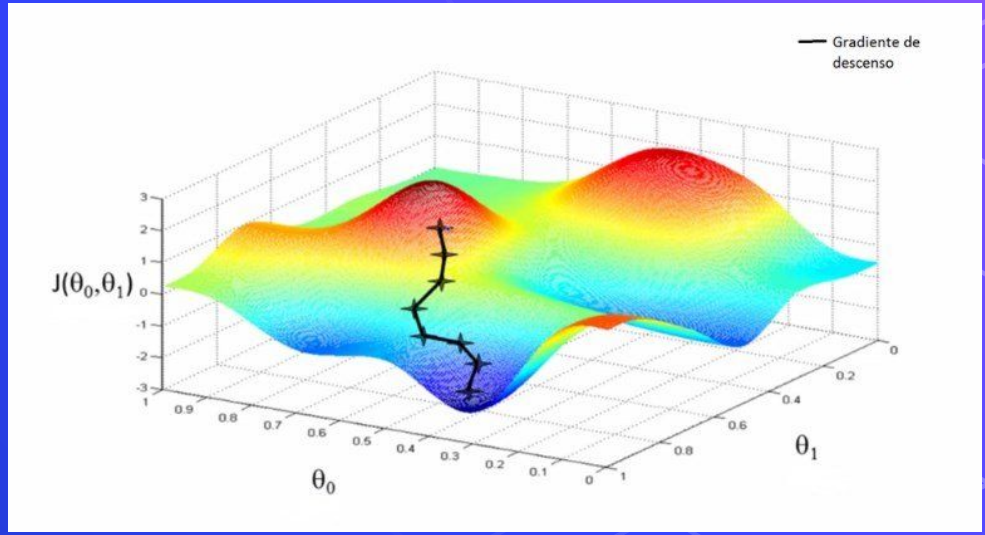
- Utilizaremos los atributos 6 (AverageMaximumTemperature) y 7 (MeanTemperature), que nos minimizan el MSE.
- Aplicaremos el algoritmo de descenso de gradiente para estos atributos y veremos el resultado gráficamente



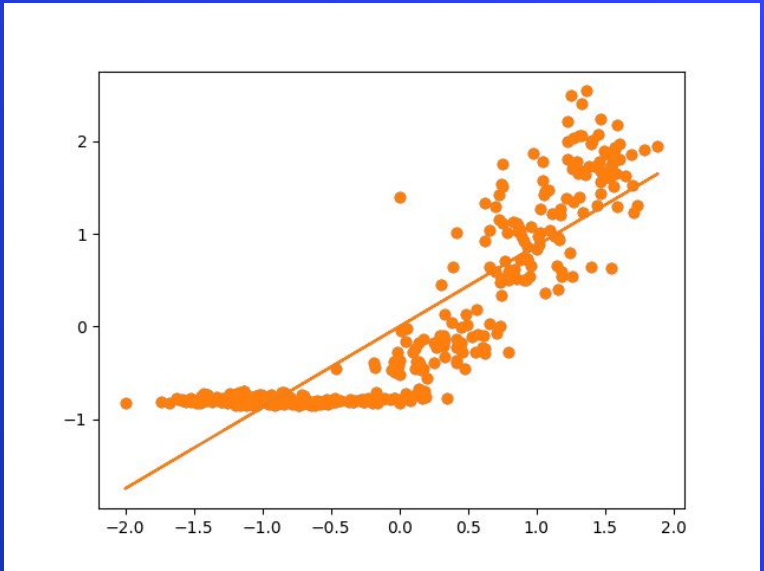
Descenso de gradiente

- Parámetros del algoritmo:
  - Alpha (aprendizaje) = 0.05
  - Iteraciones = 500

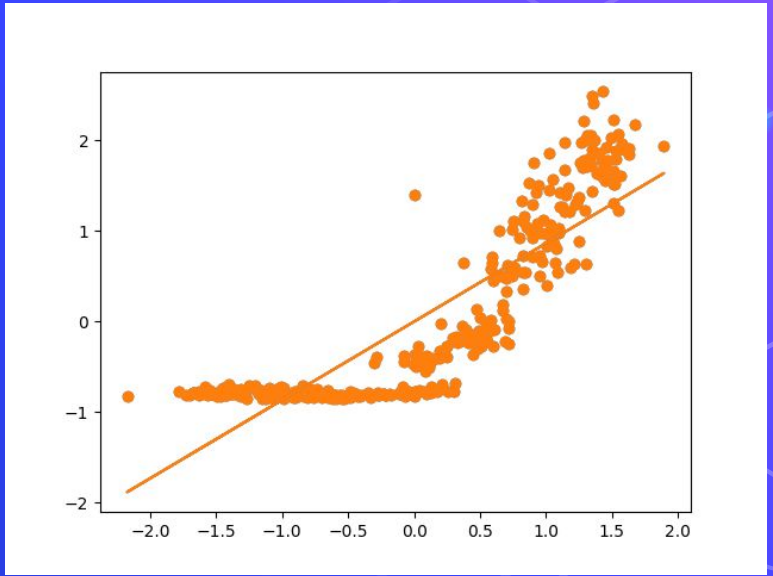
- Buena aproximación sin mucho tiempo de ejecución



Descenso de gradiente



Atributo 6 (AverageMaximumTemperature)



Atributo 7 (MeanTemperature)

Conclusiones