Michael Chan - 18562
Laboratorio 4 Data Mining

- Información del dataset
    - Se nos presenta un dataset con información de personas y basado a esto cuánto cobra una compañía de seguros por asegurarlos. Contiene información de edad, sexo, índice de masa corporal, cantidad de niños, si son fumadores, región y la cantidad que se cobra por el seguro.
- Hipótesis u objetivo
    - Ya que se tienen todos estos datos junto al cobro de seguro, se puede brindar una regresión, ya sea lineal o polinomial, con la cual podamos predecir el costo del seguro de una persona según su información.
- Solución y exploración: incluye la lógica detrás de la manipulación que realiza a los datos, retos encontrados y cualquier información relevante del proceso de exploración y delimitación del problema. Incluya gráficos y visualizaciones generadas.
    - Primero se analizan los datos:

```
In [6]: data.shape
Out[6]: (348, 7)
```

```
In [10]: data.describe()
```

Out[10]:

	age	sex	bmi	children	smoker	region	charges
count	348.000000	348.000000	348.000000	348.000000	348.000000	348.000000	348.000000
mean	39.591954	0.508621	30.676552	1.091954	0.232759	1.497126	14016.426293
std	14.417015	0.500646	5.625850	1.192021	0.423198	1.104089	12638.887852
min	18.000000	0.000000	15.960000	0.000000	0.000000	0.000000	1137.011000
25%	27.000000	0.000000	26.782500	0.000000	0.000000	1.000000	4888.466125
50%	40.000000	1.000000	30.300000	1.000000	0.000000	2.000000	9719.305250
75%	53.000000	1.000000	34.777500	2.000000	0.000000	2.000000	19006.316150
max	64.000000	1.000000	49.060000	5.000000	1.000000	3.000000	51194.559140

```
In [39]: data.head()
```

Out[39]:

	age	sex	bmi	children	smoker	region	charges
0	19	0	27.900	0	1	3	16884.92400
1	18	1	33.770	1	0	2	1725.55230
2	28	1	33.000	3	0	2	4449.46200
3	33	1	22.705	0	0	1	21984.47061
4	32	1	28.880	0	0	1	3866.85520

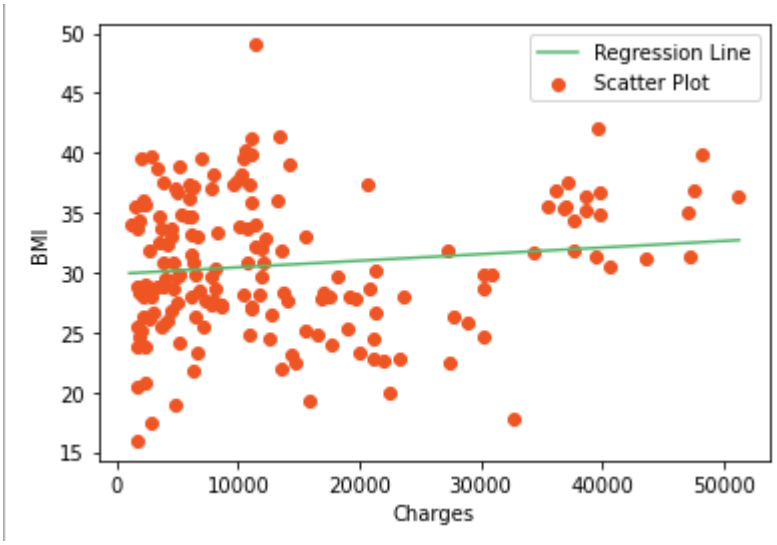
Luego de esto se revisa si existen datos que puedan a llegar a afectarnos para el análisis, como n/a o algo por el estilo.

Luego separamos nuestro dataset en la data que utilizaremos de entrenamiento y la de prueba.

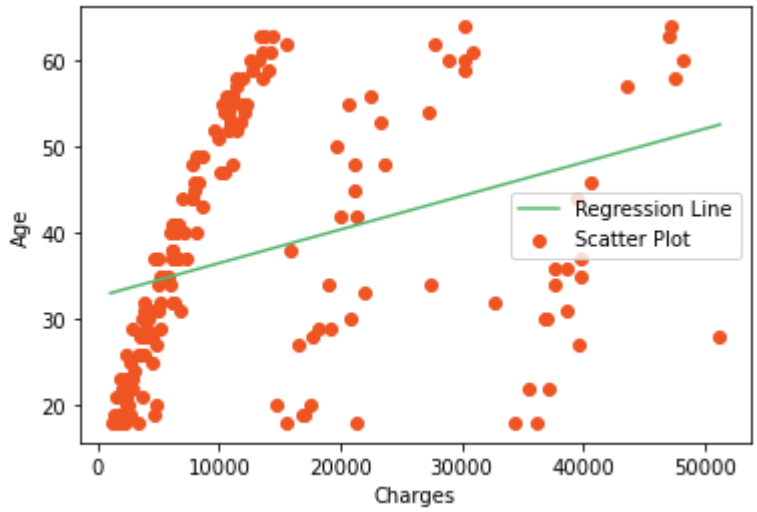
Finalmente realizaremos nuestros modelos de regresión lineal y polinomial, analizando también datos acerca de los errores comparando los datos, los cuales son MAE, MSE y RMSE

- Resultados: puntuales y respaldados por gráficos, tablas y visualizaciones en general

Regresión lineal con bmi (masa corporal):

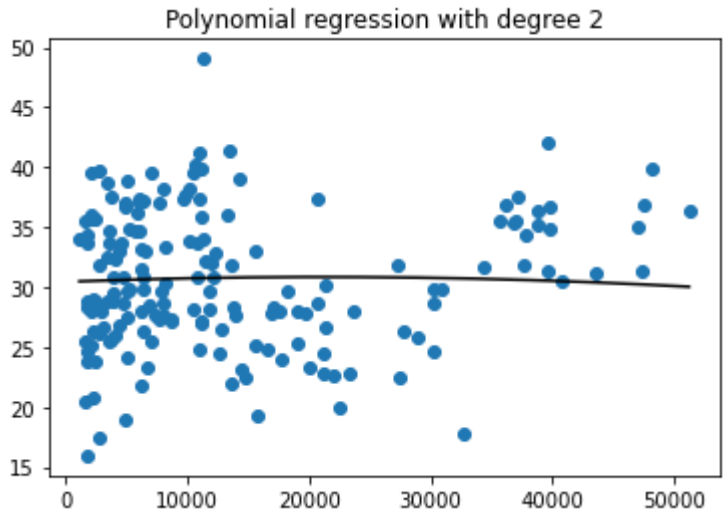


Regresión lineal con edad:



Mean absolute Error: 18.57471264367816
Mean Squared Error: 489.9080459770115
Root Mean Squared Error: 22.133866494063152

Regresión polinomial bmi:



Regresión polinomial edad: