Assignment 2

September 2, 2022

1 Assignment 2

Matthias Rathbun, Mir Khan, Jay Nagabhairu, Jason Ng, Phoebe Collins
09/09/2022

1.1 Import Libraries

```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import statsmodels
import statsmodels.api as sm
import statsmodels.formula.api as smf
import plotly.graph_objects as go
import plotly.express as px
from sklearn.metrics import mean_squared_error
from sklearn.metrics import r2_score
from sklearn.metrics import mean_absolute_percentage_error
from sklearn.metrics import mean_absolute_error
pd.set_option('display.float_format', lambda x: '%.3f' % x)
```

1.2 Problem 2.1 (0 Points) Read the CSV file "House_Prices_PRED.CSV"

```python
[2]: df = pd.read_csv("House_Prices_PRED.csv")
```

```python
[3]: df.head()
```

```
[3]:    Id  P_SalePrice  SalePrice
    0   1   206307.736      208500
    1   2   179044.533      181500
    2   3   217258.434      223500
    3   4   161547.632      140000
    4   5   272594.247      250000
```

```python
[4]: y = df["SalePrice"].values
     y_pred = df["P_SalePrice"]
```

1.3 Problem 2.2 (3 Points) Write a program to calculate the sum squared of error and the average squared error of the Model (i.e., P_House_Price).

```
[5]: sse = np.sum((y_pred - y)**2)
     print(sse)
```

740014639177.1643

```
[6]: mse = mean_squared_error(y, y_pred)
     print(mse)
```

506859341.9021673

1.4 Problem 2.3 (3 Points) Write a program to calculate the R2 of the Model (i.e., P_House_Price).

```
[7]: r2 = r2_score(y, y_pred)
     print(r2)
```

0.9196327362106914

1.5 Problem 2.4 (3 Points) Write a program to calculate the MAPE of the Model (i.e., P_House_Price).

```
[8]: mape = mean_absolute_percentage_error(y, y_pred)
     print(mape)
```

0.07026392138631052

1.6 Problem 2.5 (3 Points) Write a program to calculate the MAE of the Model (i.e., P_House_Price).

```
[9]: mae = mean_absolute_error(y, y_pred)
     print(mae)
```

12470.833673842466

1.7 Problem 2.6 (3 Points) Write a program to produce a residual plot with residual on the Y-axis and observed value (House_Price) and to impose a loess line on the graph.

```
[10]: df["residual"] = y-y_pred
      fig = px.scatter(df, x=df["SalePrice"], y=df["residual"],
                     opacity=0.8, color_discrete_sequence=['black'], trendline =␣
      ↪"lowess", trendline_options=dict(frac=0.2))

      # Change chart background color
      fig.update_layout(dict(plot_bgcolor = 'white'))
```

```python
# Update axes lines
fig.update_xaxes(showgrid=True, gridwidth=1, gridcolor='lightgrey',
                 zeroline=True, zerolinewidth=1, zerolinecolor='lightgrey',
                 showline=True, linewidth=1, linecolor='black')

fig.update_yaxes(showgrid=True, gridwidth=1, gridcolor='lightgrey',
                 zeroline=True, zerolinewidth=1, zerolinecolor='lightgrey',
                 showline=True, linewidth=1, linecolor='black')

# Set figure title
fig.update_layout(title=dict(text="Total COVID Deaths based on Total Cases",
                             font=dict(color='black')))

# Update marker size
fig.update_traces(marker=dict(size=3))
fig.update_layout(
    autosize=True,
    height=1000,)

fig.show()
```


Total COVID Deaths based on Total Cases