```python
In [93]: import pandas as pd
         import seaborn as sns
         import matplotlib.pyplot as plt
         from scipy.stats import pearsonr,spearmanr
         import warnings
         warnings.filterwarnings('ignore')
```

```python
In [94]: df = pd.read_excel('Quality_Analyst_Tolling_Data_1.xlsx') #Import clean data
         df.head()
```

Out[94]:

	License Plate	Date	Entry Time	Exit Time	Entry Toll Plaza	Exit Toll Plaza	Fee	Direction	Agency	License Plate State	Transponder Number	Transponder Agency	Facility	Vehicle Type
0	UT7ZR	2022-12-25	22:44:31	22:46:51	Plaza 2	Plaza 5	2.0	East	Agency 7	NC	8EA6UK	PQR	Facility A	Cars
1	WV30H	2022-12-14	21:54:52	22:06:35	Plaza 2	Plaza 1	4.0	West	Agency 4	LA	OHPTTW	GHI	Facility A	Cars
2	NJWWV	2022-12-12	11:12:24	11:22:57	Plaza 2	Plaza 2	2.0	West	Agency 6	NE	TUVHGT	PQR	Facility A	Buses
3	INK3P	2022-12-09	11:49:35	11:59:23	Plaza 2	Plaza 5	10.0	East	Agency 3	LA	Y8AU51	DEF	Facility A	Cars
4	NCTND	2022-12-16	07:10:05	07:10:43	Plaza 2	Plaza 2	1.0	North	Agency 3	CA	KVNFY4	JKL	Facility B	Cars

```python
In [95]: df2 = pd.read_excel('Tolling_entry_exit.xlsx')
         df2.head()
```

Out[95]:

	Entry Toll Plaza	Exit Toll Plaza	Facility	Direction	Distance
0	Plaza 1	Plaza 1	Facility A	East	3.00
1	Plaza 1	Plaza 2	Facility A	East	5.57
2	Plaza 1	Plaza 3	Facility A	East	7.88
3	Plaza 1	Plaza 4	Facility A	East	11.88
4	Plaza 1	Plaza 5	Facility A	East	15.37

```python
In [96]: df.dtypes #Checks the data types for each column
```

Out[96]:
```
License Plate              object
Date              datetime64[ns]
Entry Time                object
Exit Time                 object
Entry Toll Plaza          object
Exit Toll Plaza           object
Fee                       float64
Direction                 object
Agency                    object
License Plate State       object
Transponder Number        object
Transponder Agency        object
Facility                  object
Vehicle Type              object
dtype: object
```

```python
In [97]: df2.dtypes
```

Out[97]:
```
Entry Toll Plaza    object
Exit Toll Plaza     object
Facility            object
Direction           object
Distance            float64
dtype: object
```

```python
In [98]: df.describe() #Calculates the summary statistics of the numerical columns
```

Out[98]:

	Fee
count	24522.000000
mean	5.023897
std	3.074155
min	0.500000
25%	2.000000
50%	5.000000
75%	8.000000
max	10.000000

```python
In [99]: df['Date'] = pd.to_datetime(df['Date'])
         df['Day of Week'] =df['Date'].dt.strftime('%A')
         df.head()
```

Out[99]:

	License Plate	Date	Entry Time	Exit Time	Entry Toll Plaza	Exit Toll Plaza	Fee	Direction	Agency	License Plate State	Transponder Number	Transponder Agency	Facility	Vehicle Type	Day of Week
0	UT7ZR	2022-12-25	22:44:31	22:46:51	Plaza 2	Plaza 5	2.0	East	Agency 7	NC	8EA6UK	PQR	Facility A	Cars	Sunday
1	WV30H	2022-12-14	21:54:52	22:06:35	Plaza 2	Plaza 1	4.0	West	Agency 4	LA	OHPTTW	GHI	Facility A	Cars	Wednesday
2	NJWWV	2022-12-12	11:12:24	11:22:57	Plaza 2	Plaza 2	2.0	West	Agency 6	NE	TUVHGT	PQR	Facility A	Buses	Monday
3	INK3P	2022-12-09	11:49:35	11:59:23	Plaza 2	Plaza 5	10.0	East	Agency 3	LA	Y8AU51	DEF	Facility A	Cars	Friday
4	NCTND	2022-12-16	07:10:05	07:10:43	Plaza 2	Plaza 2	1.0	North	Agency 3	CA	KVNFY4	JKL	Facility B	Cars	Friday

```python
In [ ]:
```

```python
In [100]: #Visualization
          df['Fee'].plot(kind="hist", title="Frequency of Fees")
```

Out[100]: <AxesSubplot:title={'center':'Frequency of Fees'}, ylabel='Frequency'>


```python
In [101]: from sklearn.preprocessing import OneHotEncoder
          object_cols = ['Vehicle Type']
          OH_encoder = OneHotEncoder(sparse=False)
          OH_cols = pd.DataFrame(OH_encoder.fit_transform(df[object_cols]))
          OH_cols.index = df.index
          OH_cols.columns = OH_encoder.get_feature_names()
          df_final = df.drop(object_cols, axis=1)
          df = pd.concat([df_final, OH_cols], axis=1)

          plt.figure(figsize=(12, 6))
          sns.heatmap(df.corr(),
                      cmap='BrBG',
                      fmt='.2f',
                      linewidths=2,
                      annot=True)
```

Out[101]: <AxesSubplot:>


```python
In [102]: plt.figure(figsize=(12,6))
          sns.heatmap(df.corr(),linewidths=1)
```

Out[102]: <AxesSubplot:>


```python
In [ ]:
```

```python
In [103]: #Calculates the total revenue for each toll plaza
          df['Fee'].sum()
```

Out[103]: 123196.0

```python
In [104]: #Group data by Facility and calculate the mean
          df.groupby('Facility')['Fee'].mean()
```

Out[104]:
```
Facility
Facility A    5.018808
Facility B    5.028937
Name: Fee, dtype: float64
```

```python
In [105]: #Group data by Direction, calculates total number of transactions per group
          df.groupby('Direction').size()
```

Out[105]:
```
Direction
East     6019
North    6156
South    6164
West     6183
dtype: int64
```

```python
In [ ]:
```