Weekly Progress Report (Week 4)

Aniket Bhatia

Task Completed

The task completed during the week :

- Practiced seaborn data visualization and linear regression
- Studied about linear function
- Studied about optimization technique
- Working on the Smart City traffic pattern dataset

Challenges and Hurdles

The major challenges that are encountered during the week :

- How Linear function is implemented in data science
- To understand how linear regression works

Lessons Learned

The knowledge gained during the week :

- Introduction to machine learning
- The Loss function and Gradient Descent
- Optimization techniques
- How Hyperparameter Tuning Works

Weekly Progress

(Smart City Project)

SmartCity.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Files

..
sample_data
test.csv
train.csv

+ Code + Text

```
[2] import matplotlib.pyplot as plt

[3] df_train = pd.read_csv("train.csv")
    df_test = pd.read_csv("test.csv")

    print("Size of training set" + str(df_train.shape))
    print("Size of test set" + str(df_test.shape))
    print('\n')
    print("columns in train" + str(df_train.columns.tolist()))
    print("columns in test" + str(df_test.columns.tolist()))
    df_train.head()
```

```
Size of training set(48120, 4)
Size of test set(11808, 3)


columns in train['DateTime', 'Junction', 'Vehicles', 'ID']
columns in test['DateTime', 'Junction', 'ID']
```

	DateTime	Junction	Vehicles	ID
0	2015-11-01 00:00:00	1	15	20151101001
1	2015-11-01 01:00:00	1	13	20151101011
2	2015-11-01 02:00:00	1	10	20151101021
3	2015-11-01 03:00:00	1	7	20151101031
4	2015-11-01 04:00:00	1	9	20151101041

```
plt.figure(figsize=(10,5))
plt.title("Different Junction vehicle count")
sns.barplot(x=df_train["Junction"], y =df_train["Vehicles"])
plt.show()
```