

Лабораторная работа №2

Импортируем библиотеки

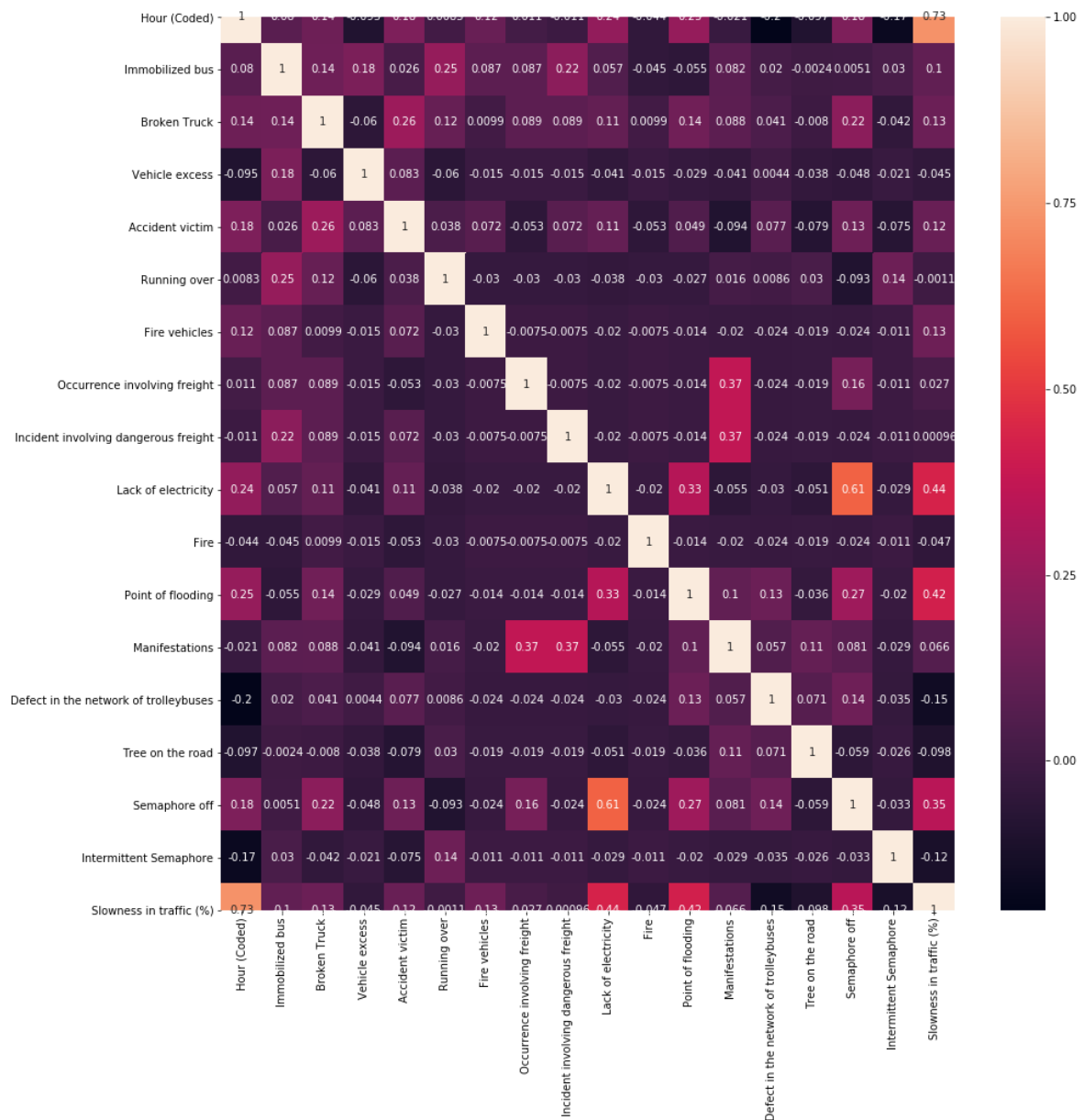
```
1 import numpy as np
2 import pandas as pd
3 from sklearn.linear_model import LinearRegression
4 from sklearn.model_selection import train_test_split
```

Затем, необходимо предоставить данные:

```
1 df = pd.read_csv('newData.csv', sep = ";")
```

Чтобы выбрать наиболее значимые переменные:

```
1 import matplotlib.pyplot as plt
2 import seaborn as sns
3
4 cols = ['Hour (Coded)', 'Immobilized bus', 'Broken Truck', 'Vehicle
5 excess',
6         'Accident victim', 'Running over', 'Fire vehicles',
7         'Occurrence involving freight', 'Incident involving dangerous
8 freight',
9         'Lack of electricity', 'Fire', 'Point of flooding',
10        'Manifestations',
11        'Defect in the network of trolleybuses', 'Tree on the road',
12        'Semaphore off', 'Intermittent semaphore', 'Slowness in traffic
13        (%)']
14
15 # figsize задает размер картинки в дюймах
16 fig, ax = plt.subplots(figsize=(15,15))
17 hm = sns.heatmap(df[cols].corr(),
18                  cbar=True,
19                  annot=True, ax=ax)
```



Исходя из этих данных:

```
1 df_2 = df[['Hour (Coded)', 'Lack of electricity', 'Point of flooding', 'Defect in the network of trolleybuses', 'Intermittent Semaphore', 'Slowness in traffic (%)']]
```

Разделяем датасет на тренировочную и выборочную:

```
1 train, test = train_test_split(df_2, test_size=0.09)
2
3 trainData = train.values
4 testData = test.values
5
6 trainX = trainData[:, :5]
7 testX = testData[:, :5]
8 trainY = trainData[:, 5:]
9 testY = testData[:, 5:]
```

Создаем модель регрессии:

```
1 model = LinearRegression().fit(trainX, trainY)
```

Получаем результаты:

```
1 from sklearn.metrics import mean_squared_error
2
3 print('Coefficient (b1): ', model.coef_)
4 print('Intercept (b0): ', model.intercept_)
5 print('Mean squared error: ', mean_squared_error(testY,
6 model.predict(testX)))
7 # Explained variance score: 1 is perfect
8 print('R² value: ', model.score(trainX, trainY))
```

Coefficient (b₁): [[0.36290307 1.83097061 1.1680219 -0.18042858 -0.30195265]]

Intercept (b₀): [4.75034036]

Mean squared error: 5.192697585302961

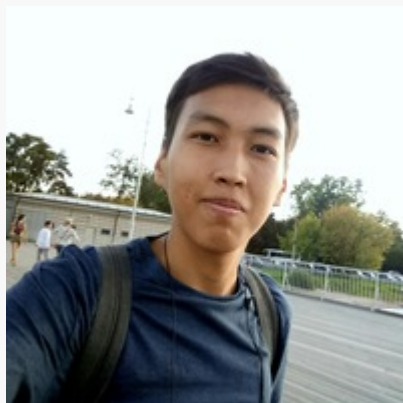
R² Value: 0.6474294781419387

```
1 y_pred = model.predict(testX)
2
3 allx = df_2.values[:, :5]
4 pred = model.predict(allx)
5 pred = pd.Series(pred.reshape(pred.shape[0],))
```

Вывод

В этой лабораторной работе мы не только работали с данными, но и анализировали их, для составления более валидных результатов. Очень интересно было создавать таблицы с matplotlib. В ходе этой работы мы научились делать правильную выборку данных.

Authors



[Arthur Kupriyanov](#)



[Artyom Kolokolov](#)

Группа: P3212