Лабораторная работа №2

Импортируем библиотеки

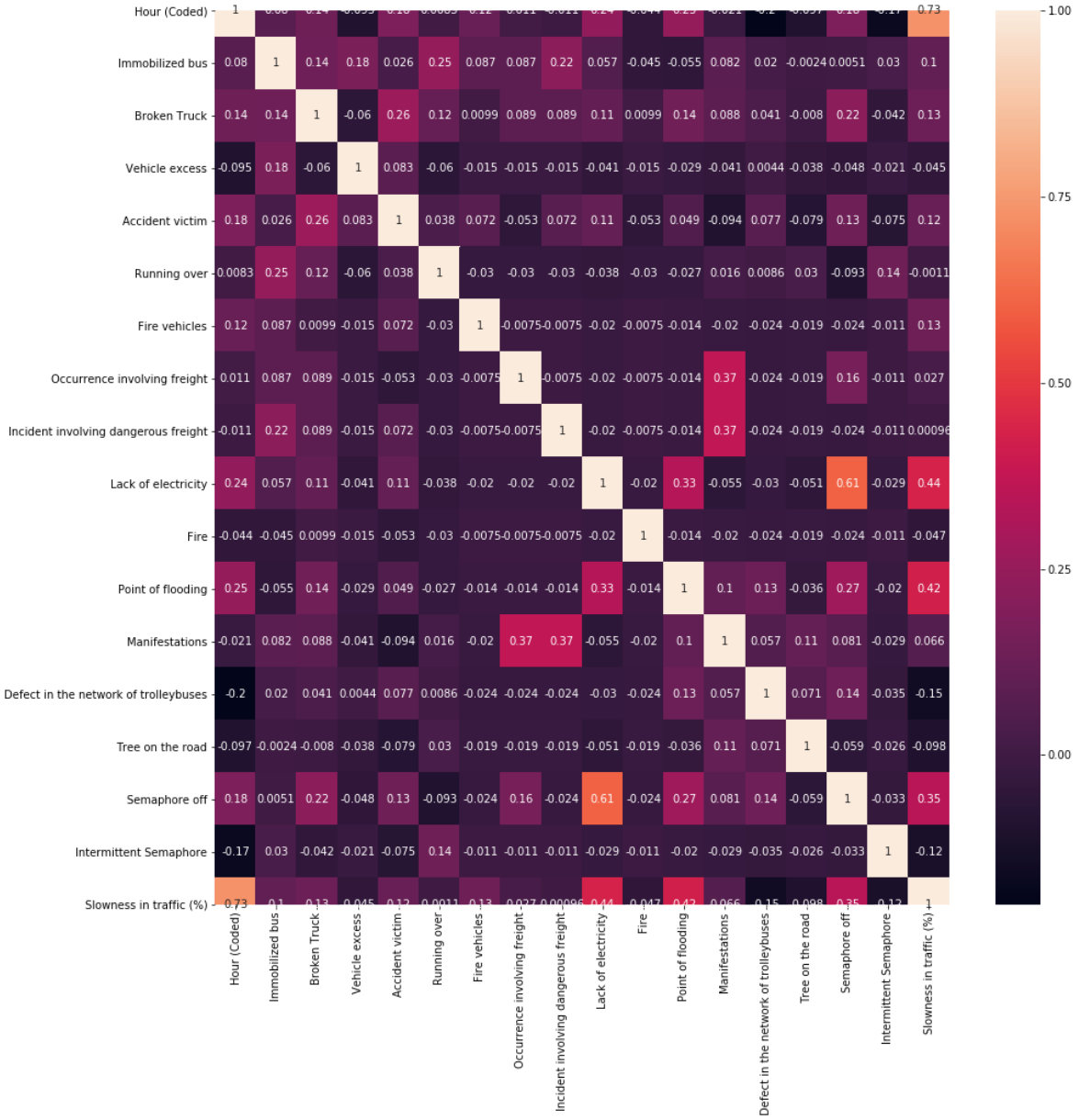
```
import numpy as np
import pandas as pd
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
```

Затем, необходимо предоставить данные:

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

Чтобы выбрать наиболее значимые переменные:

```
import matplotlib.pyplot as plt
import seaborn as sns

cols = ['Hour (Coded)', 'Immobilized bus', 'Broken Truck', 'Vehicle excess',
       'Accident victim', 'Running over', 'Fire vehicles',
       'Occurrence involving freight', 'Incident involving dangerous freight',
       'Lack of electricity', 'Fire', 'Point of flooding', 'Manifestations',
       'Defect in the network of trolleybuses', 'Tree on the road',
       'Semaphore off', 'Intermittent Semaphore', 'Slowness in traffic (%)']
# figsize задает размер картинки в дюймах
fig, ax = plt.subplots(figsize=(15,15))
hm = sns.heatmap(df[cols].corr(),
                 cbar=True,
                 annot=True, ax=ax)
```



Исходя из этих данных:

```
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:

```
train, test = train_test_split(df_2, test_size=0.09)

trainData = train.values
testData = test.values

trainX = trainData[:, :5]
testX = testData[:, :5]
trainY = trainData[:, 5:]
testY = testData[:, 5:]
```

Создаем модель регрессии:

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

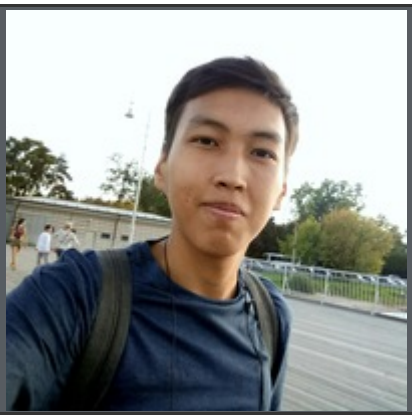
Получаем результаты:

```
from sklearn.metrics import mean_squared_error

print('Coefficient (b₁): ', model.coef_)
print('Intercept (b₀): ', model.intercept_)
print('Mean squared error: ', mean_squared_error(testY, model.predict(testX)))
# Explained variance score: 1 is perfect
print('R² Value: ', model.score(trainX, trainY))
```

> *Coefficient (b_1): [[0.36290307 1.83097061 1.1680219 -0.18042858 -0.30195265]]*
>
> *Intercept (b_0): [4.75034036]*
>
> *Mean squared error: 5.192697585302961*
>
> *R^2 Value: 0.6474294781419387*

```
y_pred = model.predict(testX)

allX = df_2.values[:, :5]
pred = model.predict(allX)
pred = pd.Series(pred.reshape(pred.shape[0],))
```

Authors

	
Arthur Kupriyanov	Artyom Kolokolov

Группа: P3212