Objectives: We would like to:

• See your skills to get insights from data.

• See how you use code in Python or R to solve this.

• Give you a chance to show your visualization, statistics and machine learning skills.

Please complete as much of the challenge as you can.

There’s no actual time limit. In the second interview, we will discuss your findings and the code.

Scenario and Data:

NYC is known for its traffic jams. So, going around on a bike is a big time saver. But there’s some risk of having an accident. We have two datasets to do some research on them:

• Bike Stations: <https://feeds.citibikenyc.com/stations/stations.json>

• Vehicle crashes:

<https://data.cityofnewyork.us/Public-Safety/NYPD-Motor-Vehicle-CollisionsCrashes/h9gi-nx95>

Questions:

• What is the most dangerous NYC borough / area for a bicycle rider? Use visualization. Done

• What would you change in the locations of bike stations to increase safety? Use visualization. Move the red coded bike station to a safer place.

• Where can an accident occur and how close is this from the nearest bike station? Create a predictive model. Geospatial Heatmap for accidents, from location and zipcode.

Time series forecast for accidents for 1 year, plot forecasted accidents map, future bike stations

• Extra questions: What other interesting patterns do you find in the data? What **other data sources** would be interesting to correlate? Population, geospatial on monthly data, per million, weather data, population information

• Extra questions: What better could you have done with respect to the current model. – LSTM models,

**Results:**

• Please share the code in an open repository (GitHub, GitLab, BitBucket, etc.; Jupyter notebook, RStudio, other?).

• Include a Readme file explaining the setup of the environment and how to run the analysis. We must be able to run the code with just the information in the Readme.

• Include a Results file describing the key insights very briefly.