BEAT

Case Study for the

Machine Learning Software Engineer position

The goal of this exercise is to model and predict the volume of passenger requests made within a specific time frame. For example, for a time frame of 1 hour, your observations should represent the number of requests per hour and your predictive model should be able to forecast future observations (the choice of time frame is up to you). Besides prediction, you are encouraged to report any significant variable interactions or correlations you may discover through your model or in a separate multivariate analysis.
You may use any software tool or programming language you deem appropriate (you may be asked to justify your choice).

Deliverables:

You should include **your source code** and a **short report** describing **your results and your approach**.

Dataset:

Dataset "`routes.csv`" contains all passenger requests along with corresponding pick-up and drop-off locations, between dates '`2015-09-01`' and '`2015-12-31`' in Lima, Peru. The file is in tab-delimited format and contains the following variables:

- `passenger_id`: unique id of every passenger
- `source_latitude`: latitude of pick-up point
- `source_longitude`: longitude of pick-up point
- `source_address`: address of pick-up point
- `destination_latitude`: latitude of destination point
- `destination_longitude`: longitude of destination point
- `destination_address`: address of destination point
- `request_date`: request timestamp (timezone Lima-Peru)

You can find the dataset for this exercise at:

https://drive.google.com/file/d/18QfvgQz5duPK6QgejP-4vfzsoWdUce_j/view?usp=sharing

Good luck! For any questions, feel free to contact `peopleops@thebeat.co`