# Traffic Management

## How may we improve the traffic congestion for Southeast Asia's roads by leveraging Grab's booking demand data?

This is a data science assignment where you are expected to create a data model from a given training dataset.

---

### PROBLEM STATEMENT

Economies in Southeast Asia are turning to AI to solve traffic congestion, which hinders mobility and economic growth. The first step in the push towards alleviating traffic congestion is to understand travel demand and travel patterns within the city.

Can we accurately forecast travel demand based on historical Grab bookings to predict areas and times with high travel demand?

### SUBMISSION DEADLINE

Please submit the final repository including documentation by or before **17 June 2019, 6.00pm (SGT)**.

**Accept the challenge**

**In this challenge, participants are to build a model trained on a historical demand dataset, that can forecast demand on a Hold-out test dataset. The model should be able to accurately forecast ahead by T+1 to T+5 time intervals (where each interval is 15-min) given all data up to time T.**

You can use the **"Demand Data"** dataset provided by Grab.

**View Datasets**

You are expected to create a Data Model based on the **"Demand Data"** dataset in order to solve the given problem statement.

You should also provide step by step documentation on how to run your code. Our evaluators will be running your data models on a test dataset.

**The given dataset contains normalised historical demand of a city, aggregated spatiotemporally within geohashes and over 15 minute intervals. The dataset spans over a two month period. A brief description of the dataset fields are found below:**

| Field | Description |
|---|---|
| geohash6 | geohash level 6<br>Geohash is a public domain geocoding system which encodes a geographic location into a short string of letters and digits with arbitrary precision. You may use the Python Geohash package **https://pypi.org/project/Geohash/** or any Java Geohash library **https://github.com/kungfoo/geohash-java** or similar to encode and decode geohash |

| | |
|---|---|
| | into latitude and longitude and vice versa. |
| day | sequential order of days |
| timestamp | start time of 15-minute intervals, in the following format: <hour>:<minute>, where hour ranges from 0 to 23 and minute is either one of (0, 15, 30, 45) |
| demand | aggregated demand normalised to be in the range [0,1] |

## You will be judged on the following criteria:

### Code Quality

Code Quality, also known as Software Quality, is generally defined in two ways:

- How well does the code conform to the functional specifications and requirements of a project.
- Structural quality, which relates to the maintainability and robustness of the code.

### Feature Engineering

Feature Engineering, also referred to as pre-processing, refers to the process of selecting and transforming variables when creating a data model for a given problem statement. While you will be given a general dataset which relates to the problem statement, you need to create “features” that make the models and algorithms work as intended.

Note that your code should be able to automatically create your desired features, that can be used in the evaluation of the Hold-out test set.

### Creativity in Problem-solving

Creativity speaks volumes about your capability to make sense of given data, derive tangible results relevant to the business needs of an organization and present the findings. All this, while keeping in mind the problem statements.

**Check out our thought process behind these challenges in our short film!**

### Model Performance

Model performance determines how a model represents the data and how well the chosen model will work. In this challenge, we will be performing a Hold-out model evaluation. For this problem, you are given a training dataset, and our evaluators will have a test dataset (not seen by the model). This test dataset will assess the likely future performance of the model.

The test dataset will cover a time period of 14 consecutive days. Submissions will be evaluated by RMSE (root mean squared error) averaged over all geohash6, 15-minute-bucket pairs.

### QUALIFICATION CRITERIA

- Submit the correct link to your repository
- Make sure your repository includes the complete codebase (all the commits are done, documentation, complete, etc)
- Solve **only one** of the challenges mentioned on the website
- Do not plagiarise the code. That will be grounds for instant disqualification

### SUBMISSION GUIDELINES

You can submit the code (either as a codebase or a Jupyter notebook) by uploading it to a **public** Github or similar repository. The instructions to submit the repository link will be sent to you via email once you accept the challenge on **https://www.aiforsea.com/**