# **Data Science Challenge - Answers**

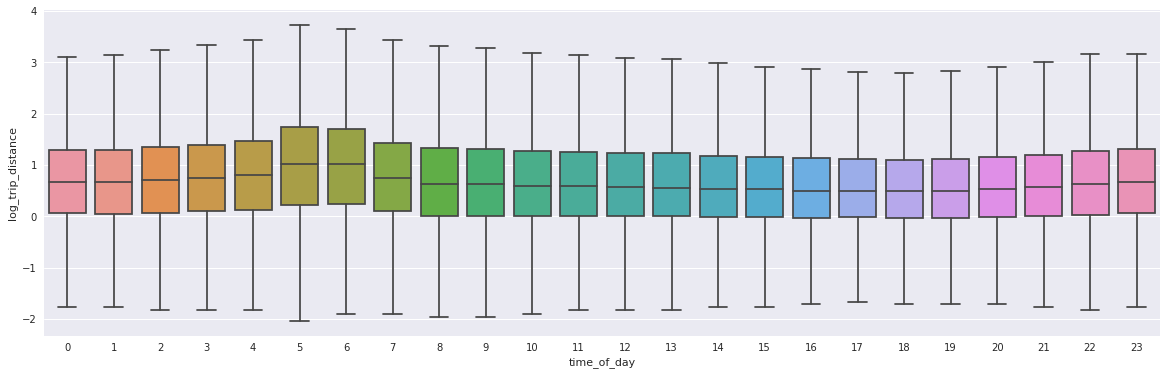
1. Download the data, load it into your favorite statistical programing software or database. Report the number of rows and columns that you've loaded.

**Answer :**

* 1. No of rows = 1,510,722
  2. No of Columns = 21

1. Visualize trip distance by time of day in any way you see fit, any observations?

**Answer** :



1. What are the most popular pickup locations on weekdays vs weekend?

**Answer:**

Weekdays

Location ID : 75

Manhattan, East Harlem South, Boro Zone

No. of pickups : 56729

Weekend

Location ID : 255

Bronx, Williamsbridge/Olinville, Boro Zone

No. of pickups : 29368

**PTO**

1. Build a model to forecast the number of trips by hour for the next 12 hours after after Feb 12th 10:00 am. How well did you do?

**Answer** : The model is very good at prediction. Average Difference between prediction and the actual trip is 17 trips

MSE on test data (x\_test,y\_test) = 253.64732828777457

**Held Out Set Performance**

Actual vs Predicted Number of Trip for April 12th from 1000 AM to 1000 PM

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Actual no. of trips** | **Predicted no. of trips** | **diff of actual vs pred** |
| **0** | 2311 | 2318.074040 | -7.074040 |
| **1** | 2175 | 2172.493378 | 2.506622 |
| **2** | 2024 | 2034.070346 | -10.070346 |
| **3** | 2106 | 2121.698199 | -15.698199 |
| **4** | 2716 | 2728.871261 | -12.871261 |
| **5** | 3009 | 3016.344247 | -7.344247 |
| **6** | 3517 | 3532.729884 | -15.729884 |
| **7** | 4040 | 4076.794780 | -36.794780 |
| **8** | 4564 | 4597.306700 | -33.306700 |
| **9** | 4725 | 4758.853024 | -33.853024 |
|  |  |  | 0 |