PUBLIC TRANSPOTATION AND ANALYSIS

PHASE 3:DEVELOPMENT PART 1

**About dataset:**

**Content:**

The public bus transportation boarding summary.csv file contains route,trip,stop and week of year from 20140711.

**Data source**

The data fields in the given file are

* **TripID** Unique identity of trip
* **RouteID** Value representing public transport route
* **StopID** Unique identity of stop
* **StopName** Name of given stop
* **WeekBeginning** Date representing first day of any week
* **NumberOfBoarding** Count of all boarding’s occurred at this stop for the named trip over the previous week

**Coding:**

In [1]:

%matplotlib inline

import numpy as np *# linear algebra*

import pandas as pd *# data processing, CSV file I/O (e.g. pd.read\_csv)*

import matplotlib.pyplot as plt

import datetime

import os

from sklearn.preprocessing import LabelEncoder

from sklearn.preprocessing import MinMaxScaler

import lightgbm as lgb

import xgboost as xgb

from sklearn.metrics import mean\_squared\_error

from math import sqrt

import warnings

warnings.filterwarnings('ignore')

print(os.listdir("../input/unisys/ptsboardingsummary"))

*# Any results you write to the current directory are saved as output.*

['Public Transport Boarding Summary by Route, Trip, Stop and Week of Year.doc', '20140711.CSV']

In [2]:

import plotly.plotly as py

import plotly.graph\_objs as go

from plotly import tools

from plotly.offline import download\_plotlyjs, init\_notebook\_mode, plot, iplot

from bubbly.bubbly import bubbleplot

init\_notebook\_mode(connected=True)

from bokeh.plotting import figure, save

from bokeh.io import output\_file, output\_notebook, show

from bokeh.models import ColumnDataSource, GMapOptions,HoverTool

from bokeh.plotting import gmap

import tensorflow as tf

from tensorflow.python.keras.models import Sequential

from tensorflow.python.keras.layers import Input, Dense, GRU,LSTM, Embedding

from tensorflow.python.keras.optimizers import RMSprop

from tensorflow.python.keras.callbacks import EarlyStopping, ModelCheckpoint, TensorBoard, ReduceLROnPlateau

In [3]:

*## For Multiple Output in single cell*

from IPython.core.interactiveshell import InteractiveShell

InteractiveShell.ast\_node\_interactivity = "all"

In [4]:

data = pd.read\_csv('../input/unisys/ptsboardingsummary/20140711.CSV')

In [5]:

out\_geo = pd.read\_csv('../input/outgeo/output\_geo.csv')

route = pd.read\_csv('../input/trann11/transit/routes.csv')

In [6]:

data.shape

data.head(2)

Out[6]:

(10857234, 6)

Out[6]:

|  | TripID | RouteID | StopID | StopName | WeekBeginning | NumberOfBoardings |
| --- | --- | --- | --- | --- | --- | --- |
| 0 | 23631 | 100 | 14156 | 181 Cross Rd | 2013-06-30 00:00:00 | 1 |
| 1 | 23631 | 100 | 14144 | 177 Cross Rd | 2013-06-30 00:00:00 | 1 |

In [7]:

route.head(2)

out\_geo.head(2)

Out[7]:

|  | route\_id | agency\_id | route\_short\_name | route\_long\_name | route\_desc | route\_type | route\_url | route\_color | route\_text\_color | RouteGroup |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 0 | 100 | 5 | 100 | Arndale Centre Interchange to Glen Osmond | via Woodville Road, Holbrooks Road, Marion Roa... | 3 | http://www.adelaidemetro.com.au/routes/100 | 0033CC | ffffff | 100-101 |
| 1 | 100B | 5 | 100B | Arndale Centre Interchange / Urrbrae Agricultu... | via Kingswood, Hawthorn, Edwardstown, North Pl... | 3 | http://www.adelaidemetro.com.au/routes/100B | 0033CC | ffffff | 100-101 |

Out[7]:

|  | accuracy | formatted\_address | google\_place\_id | input\_string | latitude | longitude | number\_of\_results | postcode | status | type |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 0 | ROOFTOP | 181 Cross Rd, Westbourne Park SA 5041, Australia | ChIJKT7I9rbPsGoRVHMHkIy-Oyk | 181 Cross Rd | -34.966656 | 138.592148 | 1 | 5041 | OK | street\_address |
| 1 | ROOFTOP | 177 Cross Rd, Westbourne Park SA 5041, Australia | ChIJ-VFZ87bPsGoRyfVgC5qbPpE | 177 Cross Rd | -34.966607 | 138.592301 | 1 | 5041 | OK | street\_address |