## Data Exploration

Total Having 1 Year of Data from date 2013-06-30 till 2014-07-06 in a Weekly interval based.

Having Total of 4165 Stops in South Australian Metropolitan Area.

In [28]:

data.nunique()

Out[28]:

TripID 39282

RouteID 619

StopID 7397

StopName 4165

WeekBeginning 54

NumberOfBoardings 400

formatted\_address 3242

latitude 3029

longitude 3008

postcode 207

type 16

route\_desc 440

dist\_from\_centre 3033

holiday\_label 3

dtype: int64

In [29]:

data.shapedata.columnsdata.head(3)

Out[29]:

(10857234, 14)

Out[29]:

Index(['TripID', 'RouteID', 'StopID', 'StopName', 'WeekBeginning',

'NumberOfBoardings', 'formatted\_address', 'latitude', 'longitude',

'postcode', 'type', 'route\_desc', 'dist\_from\_centre', 'holiday\_label'],

dtype='object')

Out[29]:

|  | TripID | RouteID | StopID | StopName | WeekBeginning | NumberOfBoardings | formatted\_address | latitude | longitude | postcode | type | route\_desc | dist\_from\_centre | holiday\_label |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 0 | 23631 | 100 | 14156 | 181 Cross Rd | 2013-06-30 | 1 | 181 Cross Rd, Westbourne Park SA 5041, Australia | -34.966656 | 138.592148 | 5041 | street\_address | via Woodville Road, Holbrooks Road, Marion Roa... | 5.180961 | 0 |
| 1 | 23631 | 100 | 14144 | 177 Cross Rd | 2013-06-30 | 1 | 177 Cross Rd, Westbourne Park SA 5041, Australia | -34.966607 | 138.592301 | 5041 | street\_address | via Woodville Road, Holbrooks Road, Marion Roa... | 5.172525 | 0 |
| 2 | 23632 | 100 | 14132 | 175 Cross Rd | 2013-06-30 | 1 | 175 Cross Rd, Westbourne Park SA 5041, Australia | -34.966758 | 138.592715 | 5041 | street\_address | via Woodville Road, Holbrooks Road, Marion Roa... | 5.180709 | 0 |

In [30]:

data.isnull().sum()

Out[30]:

TripID 0

RouteID 0

StopID 0

StopName 0

WeekBeginning 0

NumberOfBoardings 0

formatted\_address 3506

latitude 0

longitude 0

postcode 425081

type 0

route\_desc 2106618

dist\_from\_centre 0

holiday\_label 0

dtype: int64

How Many different type of Unique Data in the dataset

In [31]:

data['WeekBeginning'].unique()

Out[31]:

array([datetime.date(2013, 6, 30), datetime.date(2013, 7, 7),

datetime.date(2013, 7, 14), datetime.date(2013, 7, 21),

datetime.date(2013, 7, 28), datetime.date(2013, 8, 4),

datetime.date(2013, 8, 11), datetime.date(2013, 8, 18),

datetime.date(2013, 8, 25), datetime.date(2013, 9, 1),

datetime.date(2013, 9, 8), datetime.date(2013, 9, 15),

datetime.date(2013, 9, 22), datetime.date(2013, 9, 29),

datetime.date(2013, 10, 6), datetime.date(2013, 10, 13),

datetime.date(2013, 10, 20), datetime.date(2013, 10, 27),

datetime.date(2013, 11, 3), datetime.date(2013, 11, 10),

datetime.date(2013, 11, 17), datetime.date(2013, 11, 24),

datetime.date(2013, 12, 1), datetime.date(2013, 12, 8),

datetime.date(2013, 12, 15), datetime.date(2013, 12, 22),

datetime.date(2013, 12, 29), datetime.date(2014, 1, 5),

datetime.date(2014, 1, 12), datetime.date(2014, 1, 19),

datetime.date(2014, 1, 26), datetime.date(2014, 2, 2),

datetime.date(2014, 2, 9), datetime.date(2014, 2, 16),

datetime.date(2014, 2, 23), datetime.date(2014, 3, 2),

datetime.date(2014, 3, 9), datetime.date(2014, 3, 16),

datetime.date(2014, 3, 23), datetime.date(2014, 3, 30),

datetime.date(2014, 4, 6), datetime.date(2014, 4, 13),

datetime.date(2014, 4, 20), datetime.date(2014, 4, 27),

datetime.date(2014, 5, 4), datetime.date(2014, 5, 11),

datetime.date(2014, 5, 18), datetime.date(2014, 5, 25),

datetime.date(2014, 6, 1), datetime.date(2014, 6, 8),

datetime.date(2014, 6, 15), datetime.date(2014, 6, 22),

datetime.date(2014, 6, 29), datetime.date(2014, 7, 6)],

dtype=object)

## Data Visualization

In [32]:

*##can assign the each chart to one axes at a time*fig,axrr=plt.subplots(3,2,figsize=(18,18))

data['NumberOfBoardings'].value\_counts().sort\_index().head(20).plot.bar(ax=axrr[0][0])data['WeekBeginning'].value\_counts().plot.area(ax=axrr[0][1])data['RouteID'].value\_counts().head(20).plot.bar(ax=axrr[1][0])data['RouteID'].value\_counts().tail(20).plot.bar(ax=axrr[1][1])data['type'].value\_counts().head(5).plot.bar(ax=axrr[2][0])data['type'].value\_counts().tail(10).plot.bar(ax=axrr[2][1])

Out[32]:

<matplotlib.axes.\_subplots.AxesSubplot at 0x7f1726f9e860>

Out[32]:

<matplotlib.axes.\_subplots.AxesSubplot at 0x7f1615adbb38>

Out[32]:

<matplotlib.axes.\_subplots.AxesSubplot at 0x7f1645050f28>

Out[32]:

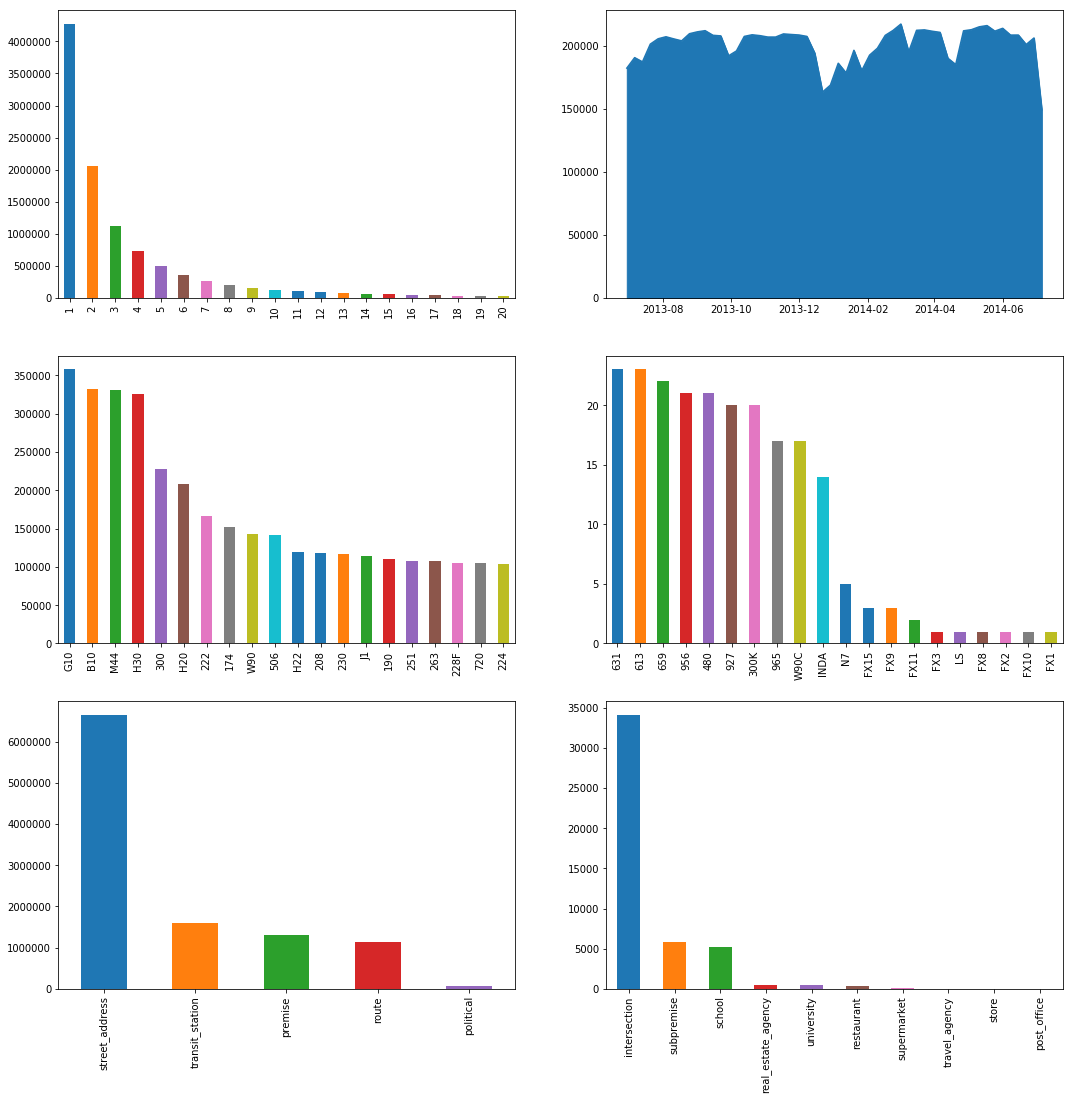
<matplotlib.axes.\_subplots.AxesSubplot at 0x7f171ef36588>

Out[32]:

<matplotlib.axes.\_subplots.AxesSubplot at 0x7f171ef5dc50>

Out[32]:

<matplotlib.axes.\_subplots.AxesSubplot at 0x7f171ef0d2e8>



****Inferences****:

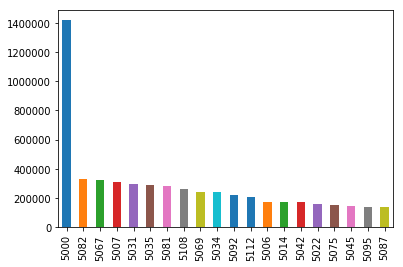
* More than 40 lakhs times only single person board from the bus stop.
* There are average of 1.8 lakhs people travel every week by bus in adelaide metropolitan area.
* G10,B10,M44,H30 are the most busiest routes in the city while FX8,FX3,FX10,FX1,FX2 are the least.
* Most of the Bus stops are Street\_Address Type while there are very few which are store or post office.

In [33]:

data['postcode'].value\_counts().head(20).plot.bar()

Out[33]:

<matplotlib.axes.\_subplots.AxesSubplot at 0x7f171b4c0c50>



In [34]:

*# data['dist\_from\_centre'].nunique()*bb\_grp = data.groupby(['dist\_from\_centre']).agg({'NumberOfBoardings': ['sum']}).reset\_index()bb\_grp.columns = bb\_grp.columns.get\_level\_values(0)bb\_grp.head()bb\_grp.columns

Out[34]:

|  | dist\_from\_centre | NumberOfBoardings |
| --- | --- | --- |
| 0 | 0.000018 | 1892443 |
| 1 | 0.131368 | 167535 |
| 2 | 0.309089 | 356518 |
| 3 | 0.314937 | 1484824 |
| 4 | 0.326005 | 120061 |

Out[34]:

Index(['dist\_from\_centre', 'NumberOfBoardings'], dtype='object')

In [35]:

trace0 = go.Scatter(

x = bb\_grp['dist\_from\_centre'],

y = bb\_grp['NumberOfBoardings'],mode = 'lines+markers',name = 'X2 King William St')

data1 = [trace0]layout = dict(title = 'Distance Vs Number of boarding',

xaxis = dict(title = 'Distance from centre'),

yaxis = dict(title = 'Number of Boardings'))fig = dict(data=data1, layout=layout)iplot(fig)

05k10k15k00.5M1M1.5M2MExport to plot.ly »Distance Vs Number of boardingDistance from centreNumber of Boardings

****Inferences****:

* As we move away from centre the number of Boarding decreases
* There are cluster of bus stops near to the main Adelaide city as oppose to outside.so that's why most of boardings are near to center

### Using Bokeh

Plot the Bus stop on the Google Map using the latitude and longitude of the bus stop address

In [36]:

lat = out\_geo['latitude'].tolist()long = out\_geo['longitude'].tolist()nam = out\_geo['input\_string'].tolist()

In [37]:

map\_options = GMapOptions(lat=-34.96, lng=138.592, map\_type="roadmap", zoom=9)key = open('../input/geolockey/api\_key.txt').read()p = gmap(key, map\_options, title="Adelaide South Australia")source = ColumnDataSource(data=dict(lat=lat,lon=long,nam=nam))

p.circle(x="lon", y="lat", size=5, fill\_color="blue", fill\_alpha=0.8, source=source)TOOLTIPS = [("Place", "@nam")]p.add\_tools( HoverTool(tooltips=TOOLTIPS))output\_notebook()show(p)

Out[37]:

|  |  |
| --- | --- |
|  |  |

****Inferences****:

* It has Geospatial coverage Area from Lat: 34.3862 to -35.3655 and Lon: 138.4126 to 139.1089. Which is Total 152 KM long Area from Daniel Road to Mosquito Creek Road on one side and Total 162 KM Stretch from Truro to Myponga Beach on the other side.
* There are cluster of bus stops near to the main Adelaide city as oppose to outside.

source\_6 = bb[bb['StopName'] == '57A Hancock Rd'].reset\_index(drop = True)source\_7 = bb[bb['StopName'] == '37 Muriel Dr'].reset\_index(drop = True)source\_8 = bb[bb['StopName'] == '18B Springbank Rd'].reset\_index(drop = True)source\_9 = bb[bb['StopName'] == '27E Sir Ross Smith Av'].reset\_index(drop = True)source\_10 = bb[bb['StopName'] == '46A Baldock Rd'].reset\_index(drop = True)

In [42]:

trace0 = go.Scatter(

x = source\_6['WeekBeginning'],

y = source\_6['NumberOfBoardings\_sum'],mode = 'lines+markers',name = '57A Hancock Rd')trace1 = go.Scatter(

x = source\_7['WeekBeginning'],

y = source\_7['NumberOfBoardings\_sum'],mode = 'lines+markers',name = '37 Muriel Dr')trace2 = go.Scatter(

x = source\_8['WeekBeginning'],

y = source\_8['NumberOfBoardings\_sum'],mode = 'lines+markers',name = '18B Springbank Rd')trace3 = go.Scatter(

x = source\_9['WeekBeginning'],

y = source\_9['NumberOfBoardings\_sum'],mode = 'lines+markers',name = '27E Sir Ross Smith Av')trace4 = go.Scatter(

x = source\_10['WeekBeginning'],

y = source\_10['NumberOfBoardings\_sum'],mode = 'lines+markers',name = '46A Baldock Rd')

data = [trace0,trace1,trace2,trace3,trace4]layout = dict(title = 'Weekly Boarding Total',

xaxis = dict(title = 'Week Number'),

yaxis = dict(title = 'Number of Boardings'),

shapes = [{*# Holidays Record: 2013-09-01*'type': 'line','x0': '2013-09-01','y0': 0,'x1': '2013-09-02','y1': 80,'line': {

'color': 'rgb(55, 128, 191)','width': 1,'dash': 'dashdot'},},

{*# 2013-10-07*'type': 'line','x0': '2013-10-07','y0': 0,'x1': '2013-10-07','y1': 80,'line': {

'color': 'rgb(55, 128, 191)','width': 1,'dash': 'dashdot'},},

{*# 2013-12-25*'type': 'line','x0': '2013-12-25','y0': 0,'x1': '2013-12-26','y1': 80,'line': {

'color': 'rgb(55, 128, 191)','width': 3,'dash': 'dashdot'},},

{*# 2014-01-27*'type': 'line','x0': '2014-01-27','y0': 0,'x1': '2014-01-28','y1': 80,'line': {

'color': 'rgb(55, 128, 191)','width': 1,'dash': 'dashdot'},},

{*# 2014-03-10*'type': 'line','x0': '2014-03-10','y0': 0,'x1': '2014-03-11','y1': 80,'line': {

'color': 'rgb(55, 128, 191)','width': 1,'dash': 'dashdot'},},

{*# 2014-04-18*'type': 'line','x0': '2014-04-18','y0': 0,'x1': '2014-04-19','y1': 80,'line': {

'color': 'rgb(55, 128, 191)','width': 3,'dash': 'dashdot'},},

{*# 2014-06-09*'type': 'line','x0': '2014-06-09','y0': 0,'x1': '2014-06-10','y1': 80,'line': {

'color': 'rgb(55, 128, 191)','width': 1,'dash': 'dashdot'},},])fig = dict(data=data, layout=layout)iplot(fig)

Jul 2013Sep 2013Nov 2013Jan 2014Mar 2014May 2014Jul 2014020406080100120Export to plot.ly »57A Hancock Rd37 Muriel Dr18B Springbank Rd27E Sir Ross Smith Av46A Baldock RdWeekly Boarding TotalWeek NumberNumber of Boardings

****Inferences****:

* Same decreasing affect of Holidays on number of people travelling through bus can be seen in other city bus stops also.
* The width of vertical blue line shows the number of holidays come within that week period.
* Two thickest blue lines shows Christmas and New year period while other one was easter & Good friday period.on both the occassion number of public holidays within week period was 3.