GoJek Technical Task Solution

Data Source:

- Click: cpc data for different provider at day level.
- Conversions: customer who clicked then paid. Clicked information stored in click data.

Data Analysis:

Following are the steps to do data analysis (solution enclosed in gojek.sql):

1. Load data from click and conversion csv to Mysql.

```
+------------+----------------+------+
| ds         | provider       | cpc  |
+------------+----------------+------+
| 2015-10-01 | TravelGuys     | 169  |
| 2015-10-07 | LalaLand       | 393  |
| 2015-10-03 | LalaLand       | 374  |
| 2015-10-01 | BobTheBuilder  | 414  |
| 2015-10-03 | TravelGuys     | 313  |
| 2015-10-09 | OlidayInn      |   0  |
| 2015-10-09 | BobTheBuilder  | 111  |
| 2015-10-02 | BobTheBuilder  | 252  |
| 2015-10-06 | LalaLand       | 314  |
| 2015-10-04 | LalaLand       | 495  |
+------------+----------------+------+
10 rows in set (0.01 sec)
```

```
+------------+----------------+------+
| ds         | provider       | gbv  |
+------------+----------------+------+
| 2015-10-11 | OlidayInn      | 3234 |
| 2015-10-05 | LalaLand       | 4957 |
| 2015-10-09 | TravelGuys     | 3321 |
| 2015-10-12 | BobTheBuilder  | 2933 |
| 2015-10-12 | BobTheBuilder  | 4309 |
| 2015-10-06 | TravelGuys     |    0 |
| 2015-10-07 | TravelGuys     | 4112 |
| 2015-10-09 | TravelGuys     | 2356 |
| 2015-10-10 | TravelGuys     | 3166 |
| 2015-10-06 | LalaLand       | 2476 |
+------------+----------------+------+
10 rows in set (0.00 sec)
```

2. Data processing i.e. calculating average cpc, average gbv and profitability (cost gbv – cost cpc) for each provider. Joined click and conversion data using ds (date) and provider to calculate the profitability.

```
+------------+----------------+----------+----------+------------+------------+----------+------------+--------+
| ds         | provider       | cpc_avg  | cpc_cost | cpc_volume | gbv_avg    | gbv_cost | gbv_volume | profit |
+------------+----------------+----------+----------+------------+------------+----------+------------+--------+
| 2015-10-01 | BobTheBuilder  | 296.0227 | 65125    | 220        | 2604.7059  | 44280    | 17         | -20845 |
| 2015-10-01 | LalaLand       | 290.1768 | 52522    | 181        | 3005.1579  | 57098    | 19         | 4576   |
| 2015-10-01 | OlidayInn      | 301.7745 | 61562    | 204        | 3119.1538  | 81098    | 26         | 19536  |
| 2015-10-01 | TravelGuys     | 302.5464 | 58694    | 194        | 3050.9286  | 85426    | 28         | 26732  |
| 2015-10-02 | BobTheBuilder  | 294.8102 | 63679    | 216        | 3623.4000  | 54351    | 15         | -9328  |
+------------+----------------+----------+----------+------------+------------+----------+------------+--------+
5 rows in set (0.00 sec)
```

3. parameterized the starting and ending date for filter data.

Data Visualization:

Following are the steps for data visualization (solution enclosed in gojek.py and gojek.ipynb)

1. Read the data (out.csv) from parameterized gojek.sql script's output.

2. Plot a bid movement chart for each provider's average cpc. Chart will vary based on starting and ending date, below chart plot from starting date "2015-10-01" to ending date "2015-10-12".
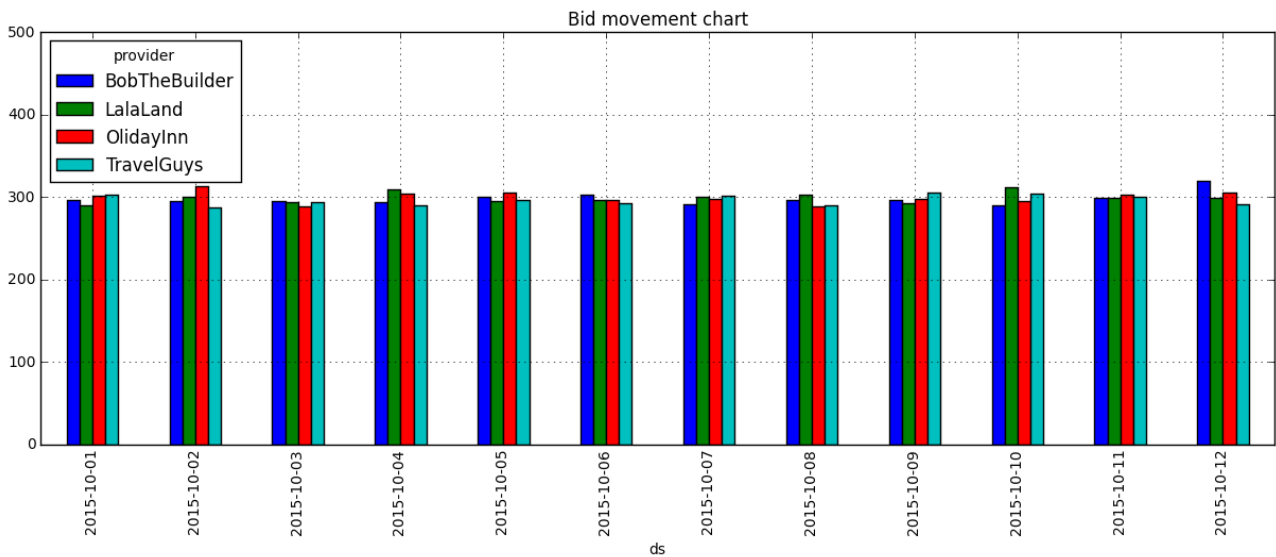
Bid movement chart

Data was generated daily (2015-10-01 to 2015-10-02) and provider's average CPC is almost same for each date i.e close to 300. To plot bid movement chart, applied the pivot function in python to transform provider information and split into multiple columns to do the comparison and plot bid chart.

3. Plot average cpc and gbv across all provider to do the comparison between average cpc and gbv.
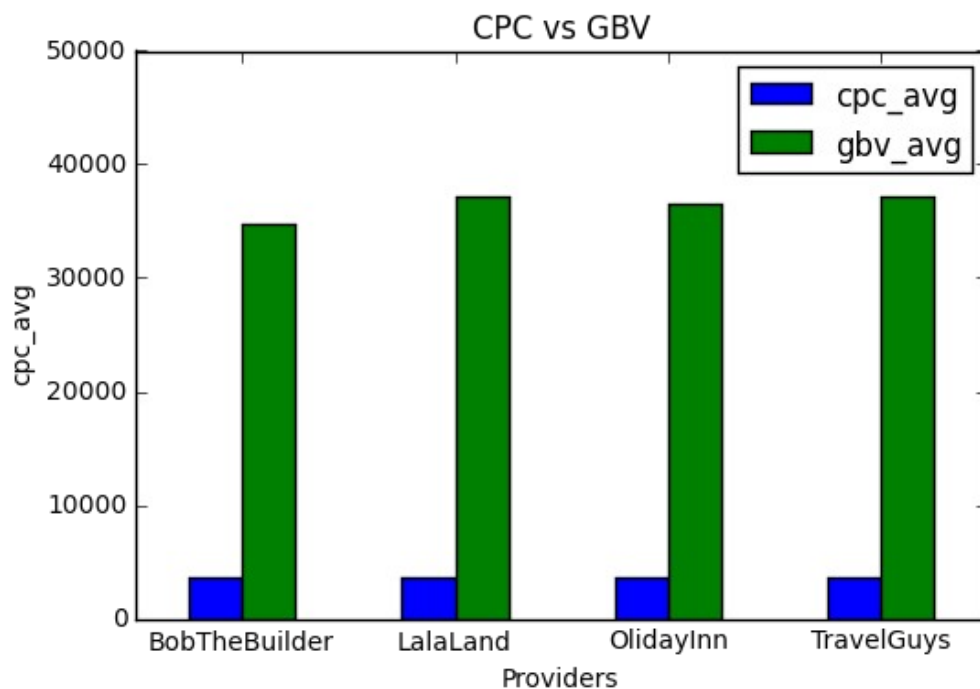
Plot bar chart to compare average cpc and gbv for each providers.

4. Analyse each provider's profitability.

Provider profitability
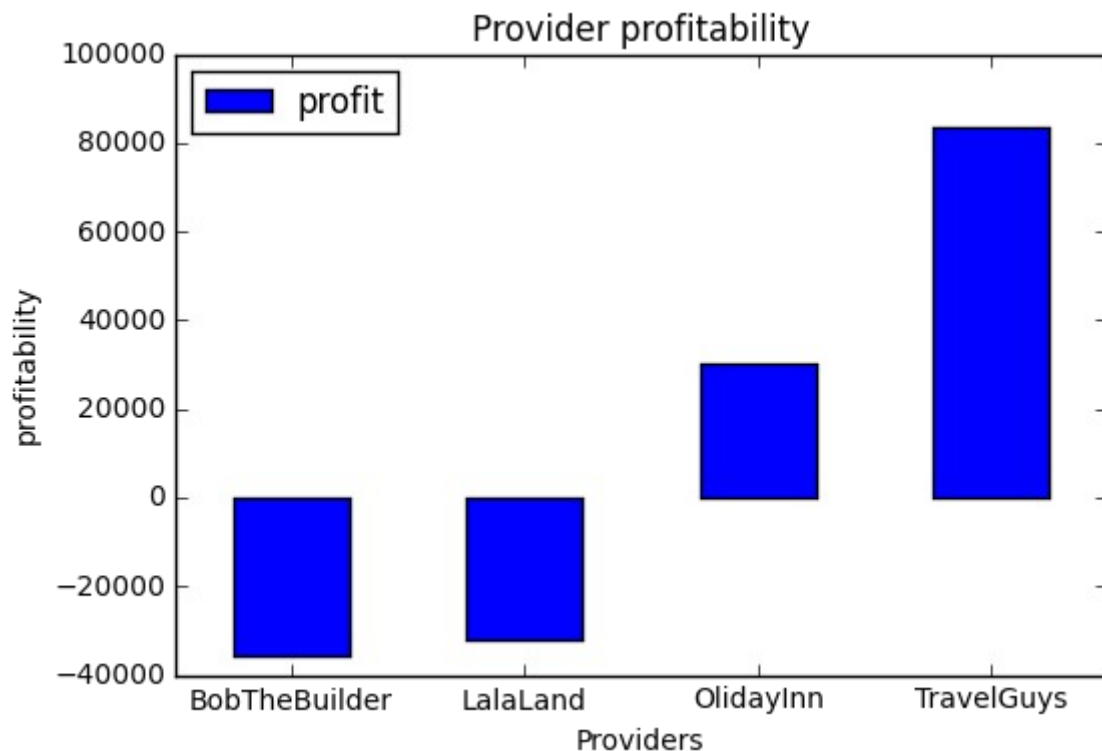
Above chart helps to analyze each provider's profitablity. TravelGuys profitablity has higest profitablity compare to other providers where as BobtheBuilder and LalaLand has negative profit (CPC cost is higher then GBV). Only TravelGuys and olidaInn provider are in profit (CPC cost is lower compare to GBV).

Bash Script excution:


```
manish@mv:~/Desktop/gojek$ sh gojek.sh gojek.sql 2015-10-01 2015-10-12
```

#run bash script with paramenters: "gojek.sql" "start date" "end date"

example: sh gojek.sh gojek.sql 2015-10-01 2015-10-12

1. gojek.sql in Bash script, creates table "clicks" and "conversions"; and load data from csv everytime.

2. Starting and Ending date parameter filters the data and store into out.csv

3. Bash script excutes python script (gojek.py), read data from out.csv and plot the graph using python matplot.