**Checklist – check figures and tables, check grammar,** #convert to pdf

1)An explanation of your approach

* 1. Any initial/high-level observations of the dataset that we’ve shared
  2. Which analytical methods have you considered applying and why

1. Any code (eg. Python, SQL, R) that you’ve used in processing/analysing the data
2. Final output
   1. The executive summary of your findings. Imagine that you would be presenting this to the management team and you need to concisely communicate the 3-5 most important insights from your analysis that could influence strategic decision making.
   2. Any charts, tables or summarised outputs that support your conclusions above
   3. Your next steps if you had more time and resources (different analytical methods, more data etc)

**Business objective**

The objective is to understand the data (traffic, transaction, onsite behavior etc.) and provide 3-5 insights for the management team.

**Data**

The data includes randomized historical data of TourRadar’s website performance from Google Analytics. It has 623076 observations with 6 variables. They include: date, path, sessions, bounces, time\_on\_page and transactions. The data dictionary is shown in table xx. There was no missing data in the provided dataset.

**Table XX: Data dictionary**

|  |  |
| --- | --- |
| **Variable** | **Variable Definition** |
| date | date of the observation |
|
| path | URL of the visit (only tour detail pages are included) |
|
| sessions | the number of the sessions |
|
| bounces | the number of bounces |
|
| time\_on\_page | the average amount of time (in seconds) users spent viewing a specified page |
|
| transactions | the number of bookings |
|

**Analytics approach**

The approach to analyzing the data would be through exploratory analysis, RFM modelling and time-series modelling. Exploratory analysis would be used to get descriptive statistics, high level observations and correlations. RFM modelling would be used to identify interesting links between path, transactions and date. Finally, time series modelling would be used to observe interesting trends over time.

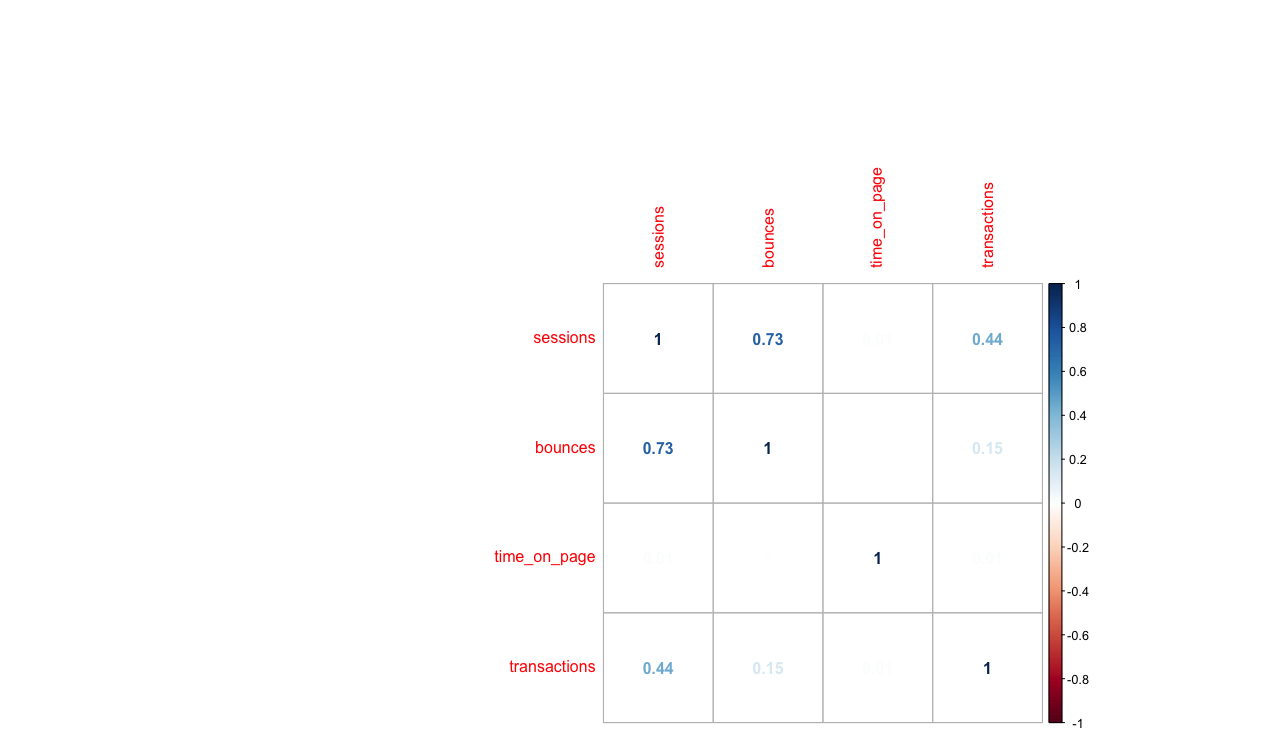
**Exploratory Analysis**

The descriptive statistics is shown in table XX. It shows the date range is between June 26, 2017 to October 1, 2017. The number of sessions ranged between 1 to 587. Someone spent more than 2 hours on a page. The largest number of transactions was 108.

**Table XX: Descriptive Statistics**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Date** | **Sessions** | **Bounces** | **Time\_on\_page** | **Transactions** |
| **Min.** | 20170626 | 1 | 0 | 0 | 0 |
| **1st Quartile** | 20170722 | 2 | 0 | 15.2 | 0 |
| **Median** | 20170815 | 4 | 0 | 61.1 | 0 |
| **Mean** | 20170811 | 4.199 | 0.2759 | 128.8 | 0.3551 |
| **3rd Quartile** | 20170909 | 6 | 0 | 148.5 | 1 |
| **Max.** | 20171001 | 587 | 250 | 7456.1 | 108 |

There seems to be a positive correlation between sessions and bounces. This is shown in Figure XX.



**Figure XX: Correlation between Sessions, Bounces, Time on page and transactions**

Looking at the information from a sessions perspective

**RFM Modelling**

**Time Series Modelling**

**Insights & Recommendations**

Based on the different analysis done here are the key takeaways.

**Next steps**

In terms of next steps, more granular data should be provided as Google analytics aggregates information. Furthermore, from the updated data it may be possible to design a machine learning model that would predict the number of transactions.