**Data Funnel Problem**

Pages:-

1. Basic Info on problem statement, Data Processing steps & Data Findings
2. Problem statement 1 to create Funnel visualization for each stage transition users
3. Problem statement 2 to Identify the TAT (Turnaround Time) across the stages.
4. Hypothesis regarding Bottleneck & It’s support analysis.

**Basic Info**

Problem Statement :-

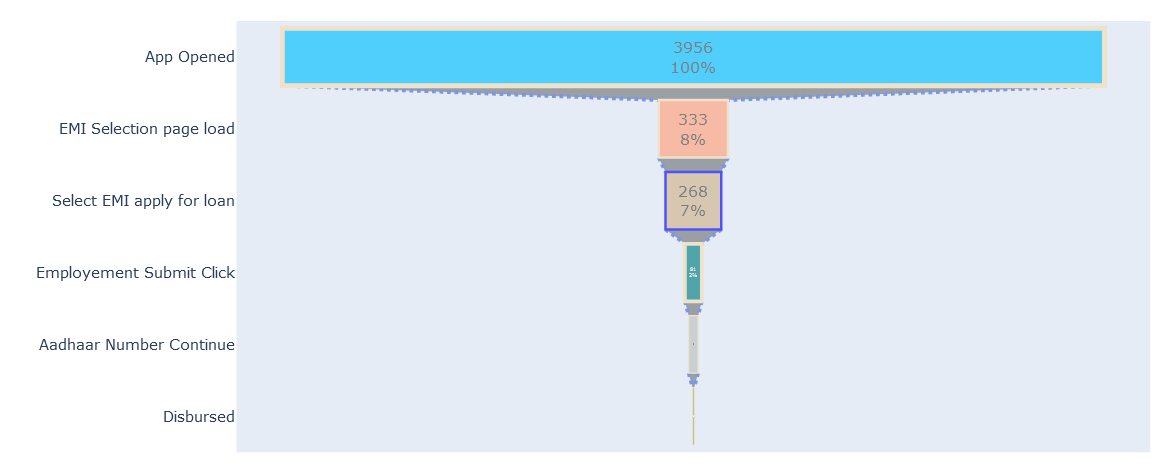
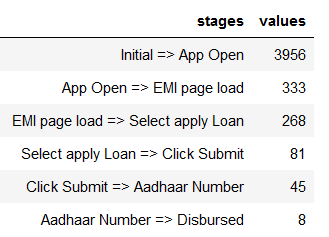
1. Create a funnel like visualization, representing the drop offs at each stage

2. Identify the TAT (Turnaround Time) across the stages (i.e time taken between each stage) in the funnel and identify the bottlenecks (if any) in the application process

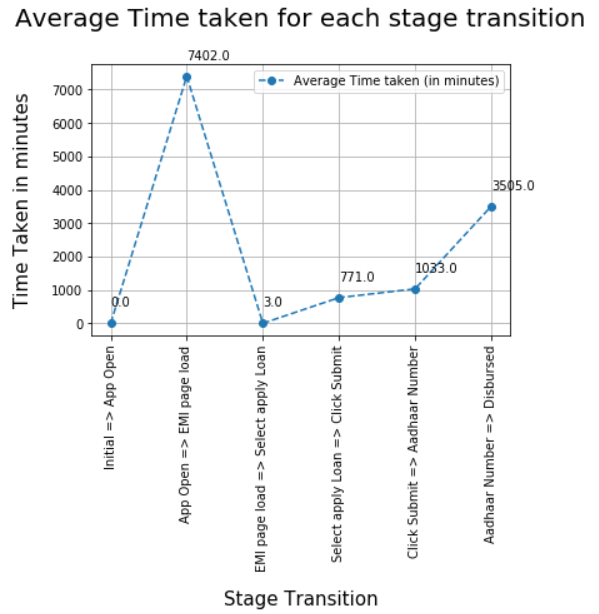
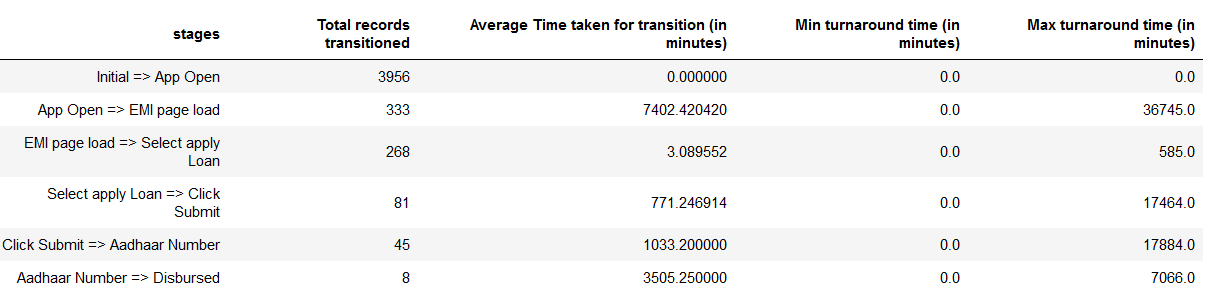
Data Processing:-

1. There are 6 excel files which consists of user\_id , event, & the timestamp that event is logged.
2. Process flow :-
   1. App\_opened -> emiselection\_page\_load -> selectemi\_applyforloan\_click -> employment\_submit\_click -> aadhaar\_number\_continue -> disbursed
3. Data Findings :-
   1. In App opened excel :-
      1. No Missing Data
      2. Timestamp both min & max timestamp is of same day “2018-09-01” varies from “00:00:58” to “15:08:46”
      3. Each user can have multiple records & it’s same for other 5 excel files.
   2. In Other excel files :-
      1. Not all user\_id are present from different file.
      2. They also can have multiple records logged for same user\_id.
      3. Time\_stamp for these files vary more than in 1 day.
4. Approach :-
   1. Select the records with minimum timestamp for each user\_id from each file. Reason as by that time our user have reached to that stage.
   2. Perform a join between each file as specified in the flow.
   3. Calculate the timestamp difference between each event for each user\_id.
   4. Sometimes timestamp difference can be negative so replace those by 0. As specified in problem statement that if second stage timestamp is less then first stage then take the last stage timestamp.
   5. This timestamp difference is termed as turnaround time for each step.
   6. Now data is ready to get the output of problem statements.

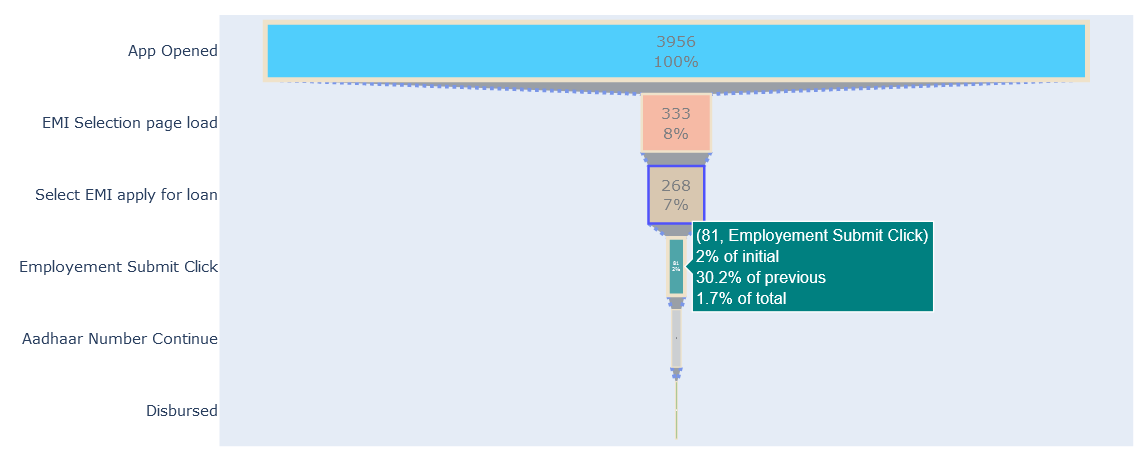
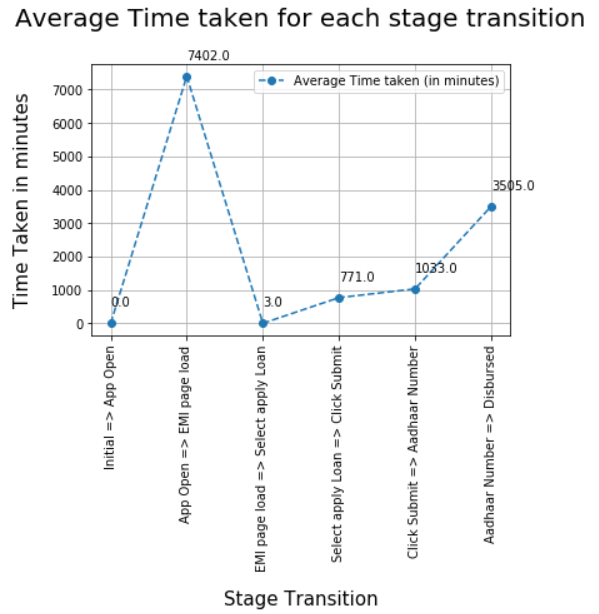
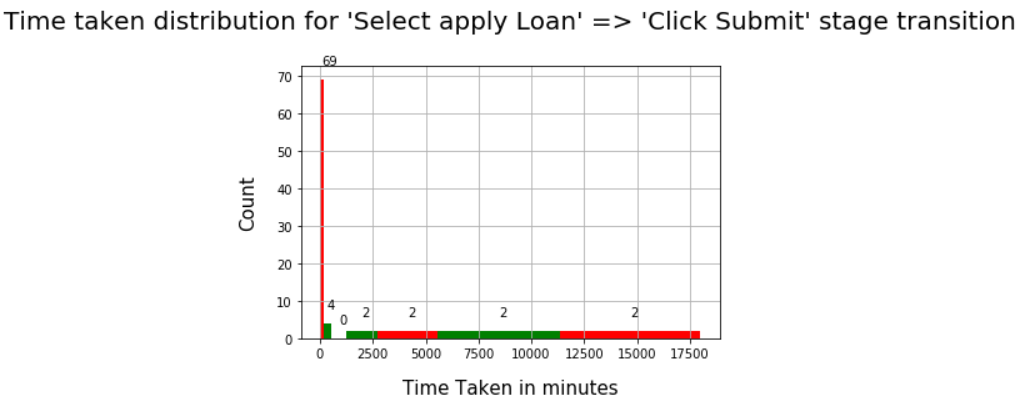
**Problem Statement 1**

* Funnel graph representing the drop offs at each stage.
* Transitioned at each step is shown in this graph that out of 3956 only 333 reached to EMI Selection page load step & so on.
* Here is the Dataframe showing the count of records transitioned at each stage.
* 

Problem Statement 2

* Identified the Turnaround time across the stages.
* For this analysis used the same dataframe transitioning which I used for the Funnel, as that will help us to better understand the transitioning between the stages for each user.
* And supporting dataframe

BottleNeck

* In Complete Transaction visible Bottleneck is at the transition of user from **selectemi\_applyforloan\_click => employment\_submit\_click.**
* Based on Funnel transitioning & Turnaround time.
* Funnel Transitioning
* As highlighted only 30.2% of user are clicking submit out of 268 users who have reached to Apply for loan stage.
* And the Turnaround Time for this transition is also high which is about 771 minutes(approx. 13 hrs) on an average which is too high for filling any form( if a form has to be filled at apply loan stage, else also it is high).
* 
* Just to Further support this Argument I have plotted the histogram plot for users who reached to this transition
* For this plot I have created a window of (3 hrs, 6hrs, 12hrs, 24hrs, 48hrs, 96hrs & rest) as shown in above graph
* There are 12 users which are taking more than 3 hrs just to pass that stage of apply loan.

**Hypothesis**

* Based on above Findings :-
* **Which means that users are facing issue while passing “selectemi\_applyforloan\_click” step & many of our users are dropping at this step that’s why our transitioning rate from this step is only 30%, we are losing 70% of users from the one who are interested in loans but due to some issues we lost them.**
* So there is need to do some changes at this step which can help our users in smooth transitioning.

**Note:-** All these findings are based on the Analysis done on **“Data Funnel.ipynb”** file.