Report

Data Analysis Assignment

At the beginning of this assignment, I downloaded the dataset and used Jupyter Notebook to preprocess the data and perform the essential operations. I have made the following observations:

Loading the lead_basics_details dataset

- All feature of dataset are categorical feature except age feature
- There is no null value in any feature of dataset
- We have outlier in age feature
- In age feature have two outlier i.e 116 and 211
- age group greater than or equel to 116 doesn't make so much effect.
- Age group 16-25 make so much effect or age 16-25 has huge contribution in dataset.
- Out of 360 row age group 16-25 available in 358 row so we conclude that age 16-25 make huge effect in dataset
- Company should more focus on this
- Female count is greater than male count.
- Lead count from chennai and Mumbai are low. Company should do something to attract lead from Chennai and Mumbai.

Loading the leads_demo_watched_details dataset

- All the feature in the dataset are categorical feature except watched_percentage

- Watched_percentage contains 2 outlier so we have to exclude it from our analysis
- In age feature have two outlier i.e 233 and 510
- watched_percentage group greater than or equal to 233 doesn't make so much effect
- watched_percentage group 2-100 make so much effect or age 2-100 has huge contribution in dataset.
- Out of 194 row age group 2-100 available in 192 row so we conclude that watch_percentage 2-100 make huge effect in dataset
- Company should more focus on this
- Now, There is no outlier in watched_percentag feature

- Lead with who has watched demo session in english and Telugu are greater in number in comparison to lead who has watched demo session in hindi .
- Very less lead wathced full demo session. Company should make shorter demo video show that lead watch full demo session.
- Most lead lead has watched demo till 90%

Loadings_interaction_details

- All the feature are categorical feature
- There is no null value in dataset
- successful call_status is 85% which is good sign for company.

- Note: We have 85% successfull call status out of this we have only 2% successfull conversation call reason so company should look at this issue.
- Most of call reason is demo_not_attended.
- We have 15% demo_schedule call reason which is a good sign for company.
- We have 70% lead as lead_stage.Project should look at conversion number

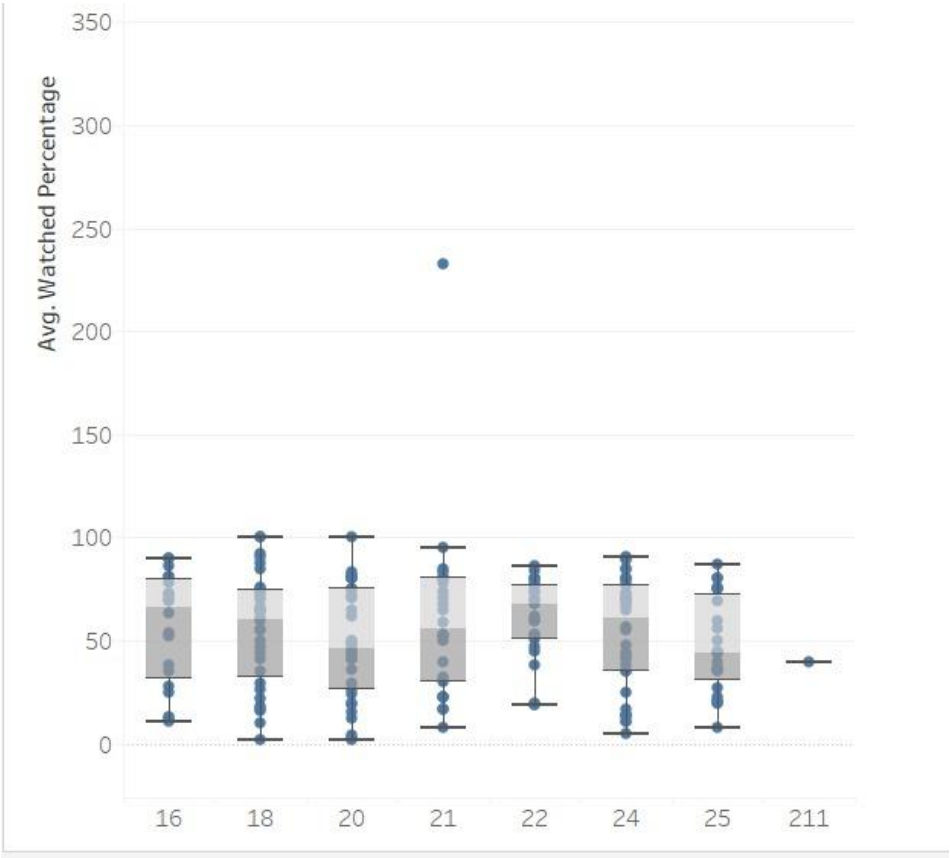
Loading the leads_reasons_for_no_interest dataset

- All the feature in the dataset are categorical feature.
- reasons_for_not_interested_in_demo has 44.2% NaN values.
- reasons_for_not_interested_to_consider hs 73.1% NaN values.
- reasons_for_not_interested_to_convert has 82.65 NaN values

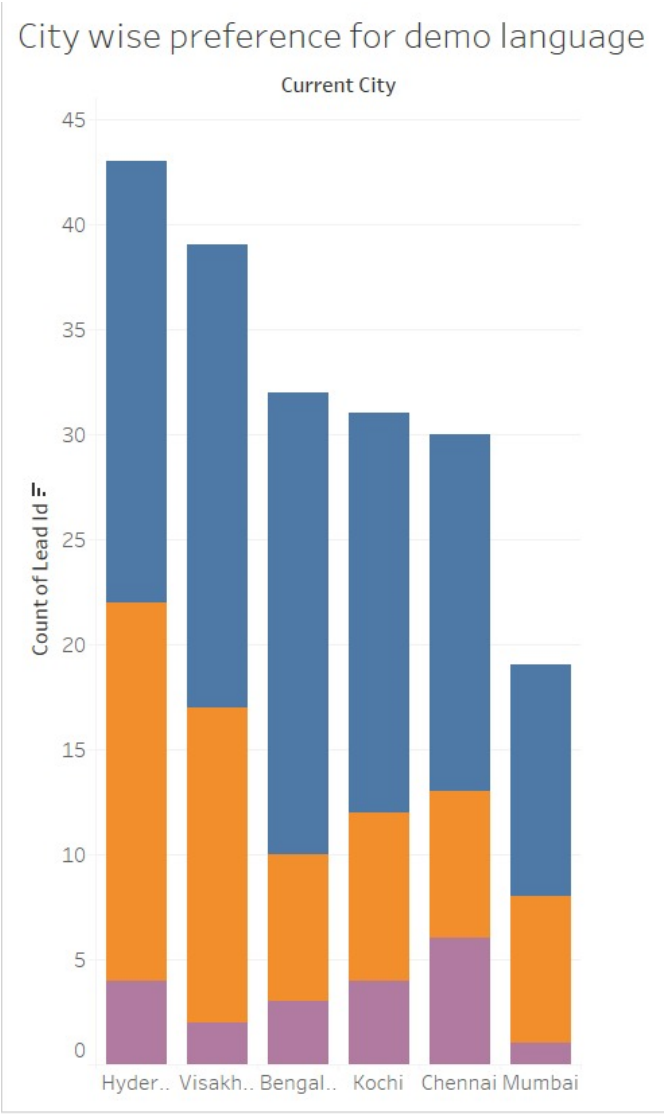
- Since all the feature are categorical feature so we can replace it with mode of particular feature but this dataset contains alot of nan values as shown above which lead wrong analysis of conclusion

Data Visualization using tableau

1.After visualizing leads and demo data in Tableau, I came across a few outliers. I have attached link of my Tableau Public account where you can see the chart:https://public.tableau.com/app/profile/shirish.waghmode/viz/GraphedOutliers/Sheet1?publish=yes

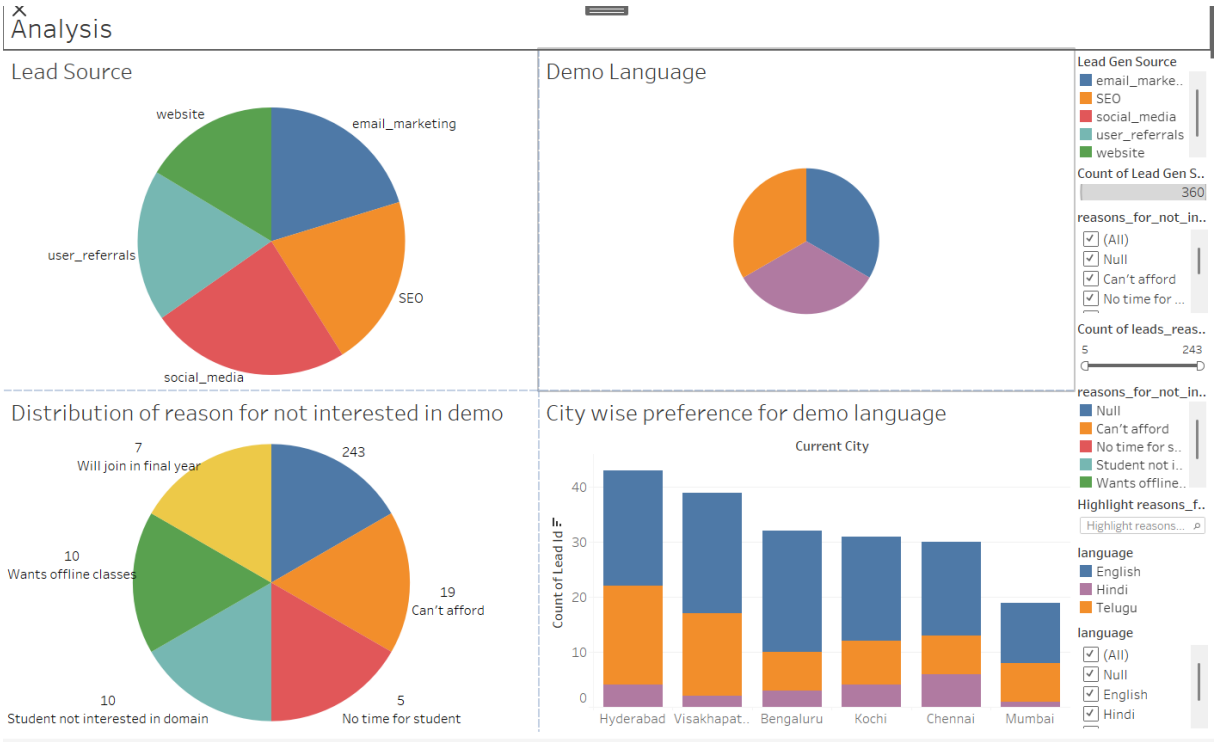


2.Here, I have performed data cleaning to move forward with the analysis
Telugu seems to be more popular in Hyderabad and Visakhapatnam, which comes under the region of Telugu-speaking people.
https://public.tableau.com/app/profile/shirish.waghmode/viz/Citywisepreference/Sheet2?publish=yes



Dashboard for Lead, language, distribution analysis

https://public.tableau.com/app/profile/shirish.waghmode/viz/DashboardforLeadlanguagedistributionanalysis/Dashboard2?publish=yes



Thank you

Submitted by-Shirish Waghmode