Visualization Product Development Report

Accommodation & Housing Trends

San Diego

Avinash Pasupulate

+91 99 4141 44 22

avinash.pasupulate@gmail.com

Table of Contents

1. Scope

The scope of this project is to create a visualization product from the housing and accommodation data collected from the San Diego region in the state of California, United States of America. The provided data would be processed and required transformations would be made to help building the visualization product to gain insights on the current trends.

2. Tools & Technologies

The provided data would be processed using either of the three tools mentioned below as required,

2.1. Python using the Pandas and other relevant packages
2.2. Tableau Desktop Public / Tableau Public Server
2.3. Microsoft Excel (Power Query)
2.4. Power Bi (if required)

3. Business Case / Requirements

The visualizations developed in the product would be used gain insights from the current trend in the housing and accommodation industry in San Diego, CA, USA

All primary datasets provided by the client would be used in developing the visualization product.

4. Queries

4.1. Business Questions

4.1.1. What are base business questions?

Define questions based on reasoning and research, since this project is not based on a client requirement.

4.2 Data Questions

4.2.1. Is there a data dictionary?

No, get through research using column names

4.2.2. One of the neighborhood names is in arabic?

This was translated to San Diego

4.2.3. There were several variations of some values in different variables

There were mapped to their right spellings / variations, not replaced in the raw data since they may occur in the subsequent updates.

5. Assumptions

Assuming that the dashboard would be used by different levels of consumers.

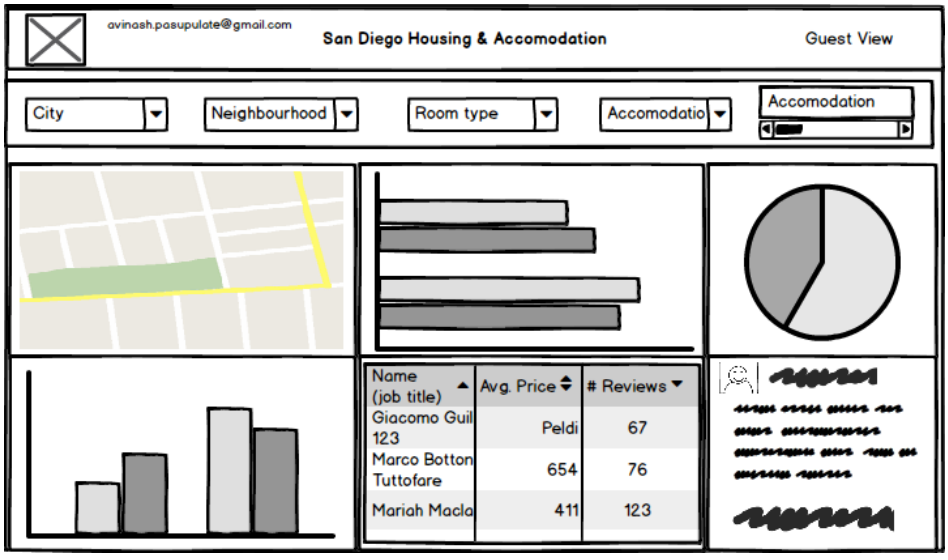
6. Consumers

The dashboards are created considering consumption across three levels of consumers,

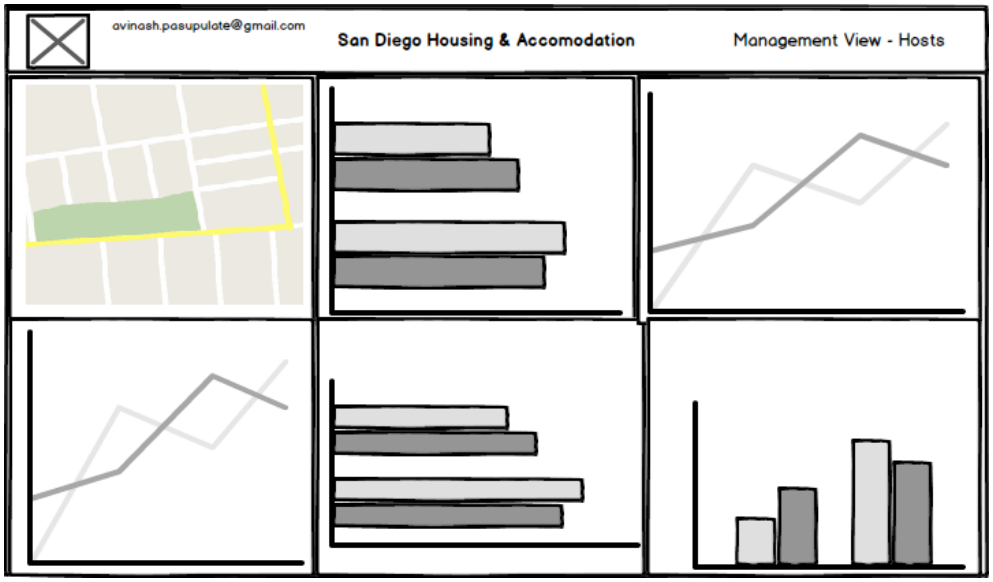
The consumers would explore the data on the created dashboards to gain insights from the data and make relevant decisions. So, three views were created for the consumers.

7. Mockups / Wireframe

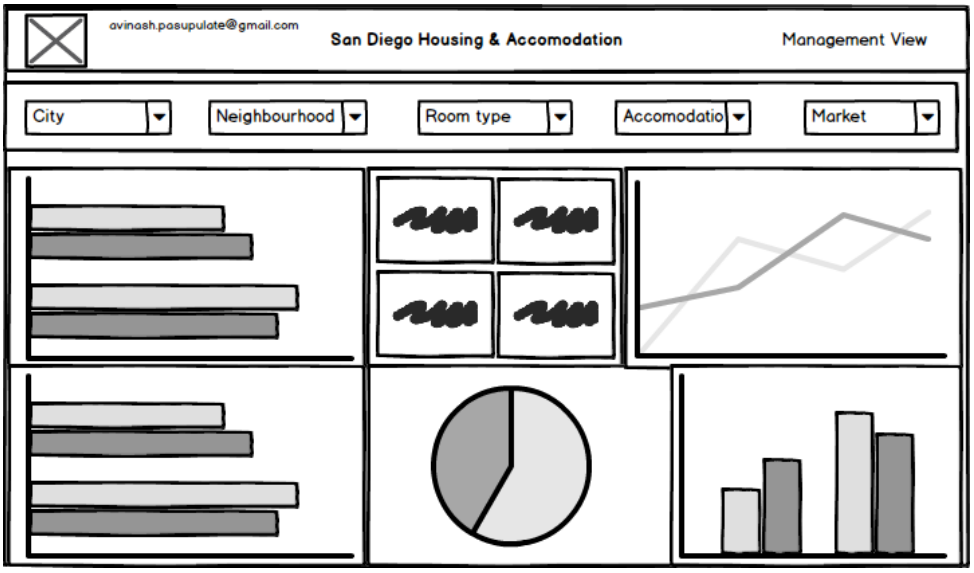
7.1. Guest View



7.2. Host View



7.3. Management View



8. Data Sources & Transformations Applied

8.1 Primary Data Sources

8.1.1 Calendar Data - calendar.csv

Calendar of available listings for a future period

8.1.2 Listings Data – listings.csv

Listing related details

8.1.3 Neighborhood Data – neighborhood.csv

Neighborhood in which the listing property is present

8.1.4 Reviews – reviews.csv

Reviews for the listings available

8.2 Secondary Data Sources

8.2.1 Sample listings – listings_1.csv

Sample listing data for qc

8.2.2 Sample reviews – reviews_1.csv

Sample reviews for qc

There were different transformations that were applied on the data, they are listed below,

8.3. Remove NA / Missing values, treatment of missing values needs to be confirmed with the client

8.4. Replace values with mapped value to remove variations in same strings (e.g. similar values in neighborhood, city etc.)

8.5. Stages for data transformation are listed in the attached Jupyter notebook

9. Quality Control

Quality control needs to be performed in a through manner since it is the most critical stage of a data product development.

9.1. Different levels (usually 2-3) needs to be performed by a 3rd person to identify deviations

9.2. Understand data sources

9.3. Create test cases

9.4. Process data and extract information to compare with the dashboard metrics

9.5. Complete test cases & list down deviations

9.6. Find root cause of the deviations and perform corrective actions

10. Deliverable / Work Product

10.1. Latest Deliverable

10.1.1. Delivering draft of the visualization product for client review and feedback

10.1.2. Report on development of the visualization product

10.1.3. Annexure documents which were used in the development of the product

10.2. Project Plan

S .No	Task	12/03/2018	13/03/2018	14/03/2018	15/03/2018	16/03/2018	17/03/2018	18/03/2018
1	Understanding the data & tasks provided							
2	Planning on the approach and generatings relevant documents							
3	Preparing guest view & making required transformations							
4	Preparing guest view with visual and descriptive changes & completion							
5	Preparing management view							
6	Prepearing development document and making visual changes to the dashboard & completion of guest & management views							
7	Preparing management dashboard, host dashboard and performing final changes							
8	Quality control & delivery of draft for client feedback							

Milestone Deliverable

11. Conclusion

Although the tool has been built further processing is required to create a final client ready report ready for delivery, Below are some points which are to be included in the future updates,

11.1. Include tertiary macro variables data to study its impact on the housing & accommodation trends in the San Diego region

11.2. Conduct a more thorough quality control

11.3. Explore on inclusion of text processing from reviews and description

11.4. Explore data and gain more insights

11.5. Further information provided in the attached Jupyter Notebook

12. Annexure

12.1. Jupyter Notebook – Data Processing

12.2. Tableau Public – Visualization

12.1. Power Bi .pbix File – Relationship created (In progress)

Version Control

S. No	Date	Clause	Revision	Reason
1	18/02/2019	NA	NA	New Document
2				
3				