DATA MANAGEMENT PROJECT REPORT

(Project Semester August-December 2022)

House Rent Prediction

Submitted by

Name : Amandeep Bareth

Registration No. : 12010838

Programme : B.Tech CSE

Section : KM006

Course Code : INT 217

Under the Guidance of

Veerpal Kaur

25909

(Faculty)

Discipline of CSE/IT

Lovely School of Computer Science and Engineering

Lovely Professional University, Phagwara

DECLARATION

I, **Amandeep Bareth** student of Computer Science and Engineering under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date : 8 November 2022 Signature :

Registration No. : **12010838** Name of the Student : **Amandeep Bareth**

ACKNOWLEDGEMENT

I would like to take this opportunity to show my sincere gratitude towards, Ms.Veerpal Kaur, our subject teacher for course code INT217 as well as our university: Lovely Professional University whose constant help and guidance allowed me to complete this project in due time. I would also like to thank my classmates whose support was crucial in helping me finish this project.

TABLE OF CONTENTS

Content	Page No.
Introduction	6
Objectives	7
Source of Dataset	8-9
ETL Process	10
Dashboard and Screenshot	11-16
Bibliography	17

INTRODUCTION

Housing in India varies from palaces of erstwhile maharajas to modern apartment buildings in big cities to tiny huts in far-flung villages. There has been tremendous growth in India's housing sector as incomes have risen. The Human Rights Measurement Initiative finds that India is doing 60.9% of what should be possible at its level of income for the right to housing.

Renting, also known as hiring or letting, is an agreement where a payment is made for the temporary use of a good, service, or property owned by another. A gross lease is when the tenant pays a flat rental amount and the landlord pays for all property charges regularly incurred by the ownership. Renting can be an example of the sharing economy.

OBJECTIVE

Objective 1 : Number of Area Locality in Specific City

The scope of analysis is to find the number of area locality in particular city by using count function. Representation can be done by using pivot table and vertical bar chart.

Objective 2 : Total Rent City Wise

This analysis can show the rent generated by specific city so that the people can get idea about the rent of the city and its expenses. This can be done by using sum function in pivot table and visualization by bar chart.

Objective 3 : Tenants Preferred in Specific City

This analysis helps people to know which city preferred which type of tenants (Bachelors or family or family/bachelors) more. This can be done by counting area locality according to the tenants types in specific city. Visualization through bar graph.

Objective 4 : Furnishing Status in Specific City

This analysis helps in getting idea about the accommodation facilities of a particular city. There is slicer in dashboard through which you can select city to get furnishing status clearly. Visualization through doughnut chart.

Objective 5 : High Rent and Low Rent Area Locality City in Specific City

This analysis helps to know the 5 area locality which are of high and low rent in particular city. Visualization through line chart.

SOURCE OF DATASET

Kaggle is an Airbnb for Data Scientists –A subsidiary of Google LLC, is an online community of data scientists and machine learning practitioners. Kaggle allows users to find and publish data sets, explore, and build models in a web-based data-science environment, work with other data scientists and machine learning engineers, and enter competitions to solve data science challenges. Kaggle got its start in 2010 by offering machine learning competitions and now also offers a public data platform, a cloud-based workbench for data science, and Artificial Intelligence education. this is where they spend their nights and weekends. It's a crowd-sourced platform to attract, nurture, train, and challenge data scientists from all around the world to solve data science, machine learning and predictive analytics problems. It has over 536,000 active members from 194 countries, and it receives close to 150,000 submissions per month. Started from Melbourne, Australia Kaggle moved to Silicon Valley in 2011, raised some 11 million dollars, then ultimately been acquired by the Google in March of 2017. Kaggle is the number one stop for data science enthusiasts all around the world who compete for prizes and boost their Kaggle rankings. There are only 94 Kaggle Grandmasters in the world to this date.

Kaggle enables data scientists and other developers to engage in running machine learning contests, write and share code, and to host datasets. The types of data science problems posted on Kaggle can be anything from attempting to predict cancer occurrence by examining patient records to analyzing sentiment to evoke by movie reviews and how this affects audience reaction. Different sources post projects on this trailblazing platform. While some are just for educational purposes and fun brain exercises, others are genuine issues that companies are trying to solve. Kaggle makes the environment competitive by awarding prizes and rankings for winners and participants. The prizes are not only monetary but can also include attractive rewards such as jobs or free products from the company hosting the competition.

Why is Kaggle Worth Your Time?

1. Interesting and challenging projects where contributors can learn and practice Kaggle competitions involve solving challenging and interesting problems. Companies post projects to numerous contributors. It especially a great place for beginners who are just trying to break into the data science field. Aside from the competitions that are open to the general public, Kaggle also has private competitions which are only open to top rated participants (Kaggle Masters).

2. Insightful discussions with Industry leaders and learned experts apart from the projects, forums are very interesting, stimulating, and informative. Through these discussions, you can either seek advice from others or offer advice to people who are dealing with issues you understand

3. Kaggle offers its audience a chance to get into the biggest data science community in the world. This platform is trusted by some of the largest data science companies of the world such as Walmart, Facebook, and Winton Capital. On Kaggle, data scientists get exposure and a chance to work on problems faced by big companies in real-time. While it is not a guarantee, there is always the chance that the company will be impressed enough to recruit.

How Kaggle Works

The host of the competition is in-charge of preparing the data and preparing a detailed description of the problem at hand. To make it more convenient for hosts, Kaggle offers an additional consulting service that can help prepare data and describe the problem in the best possible format.

Do Kaggle Projects have any Real impact

One of its biggest and most recognized projects is one by Heritage Health which offered a remarkable cash price of $3 million. Competitions hosted on Kaggle have had far-reaching impacts such as enhancing and enabling state of the art HIV/AIDS research and improving traffic forecasting. Essentially, Kaggle has given companies the opportunity to seek solutions from the best data scientist in the world and to have external pairs of eyes to look at the problems they are trying to solve.

House Rent Prediction Dataset was uploaded by Sourav Banerjee on www.kaggle.com in Aug 2022.

In this Dataset, we have information on almost 4700+ Houses/Apartments/Flats Available for Rent with different parameters like BHK, Rent, Size, No. of Floors, Area Type, Area Locality, City, Furnishing Status, Type of Tenant Preferred, No. of Bathrooms, Point of Contact.

This Dataset is created from https://www.magicbricks.com/.

Link : https://www.kaggle.com/datasets/iamsouravbanerjee/house-rent-prediction-dataset?q=rent

ETL

ETL stands for Extract, Transform and Load and it is the process of extracting data from a data source, transforming it using specialized software and then loading the transformed data into a data warehouse.

The data of House Rent Prediction Dataset, extracted from Kaggle, which is an open- source database website. The data is then transformed (cleaned) using Excel tools and Tableau then saved to the device. The dataset has 11 columns related to the election like, name of candidates and total vote polled etc. Following are the operations performed in Tableau to transform the data:

- First, fetch the dataset House Rent Prediction Dataset from Kaggle.
- Analyze the dataset carefully to look for any ambiguity.
- Load the dataset into Tableau Prep and perform the cleaning operation.
- Take the output of the cleaned dataset.
- Change the output from '. hyper' to '.csv' and then run the flow.
- Once the flow is completed the cleaned dataset is saved in the computer.

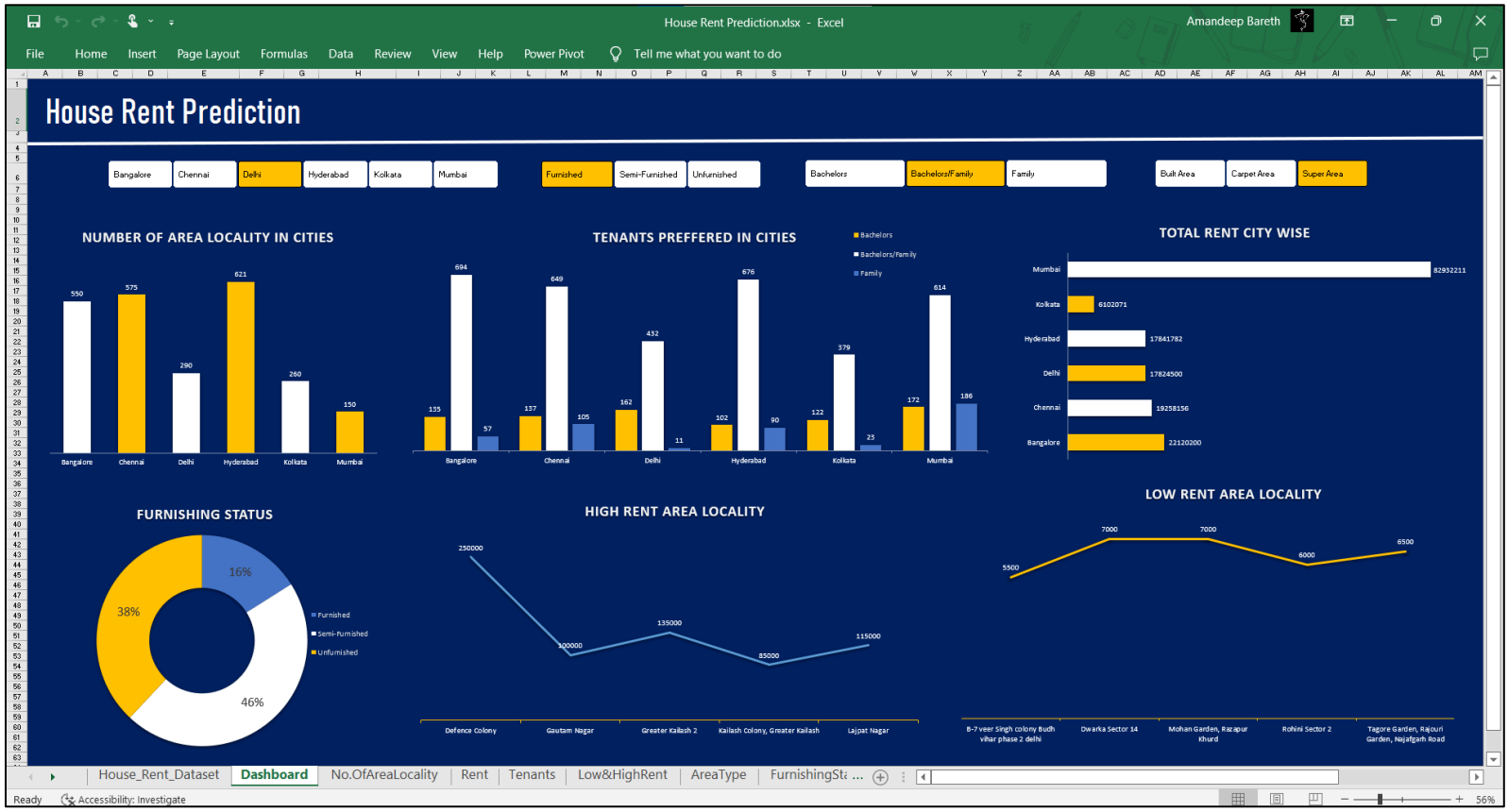
Why Do We Need ETL?

It is essential to properly format and prepare data to load it in the data storage system of your choice. The triple combination of ETL provides crucial functions that are many times combined into a single application or suite of tools that help in the following areas:

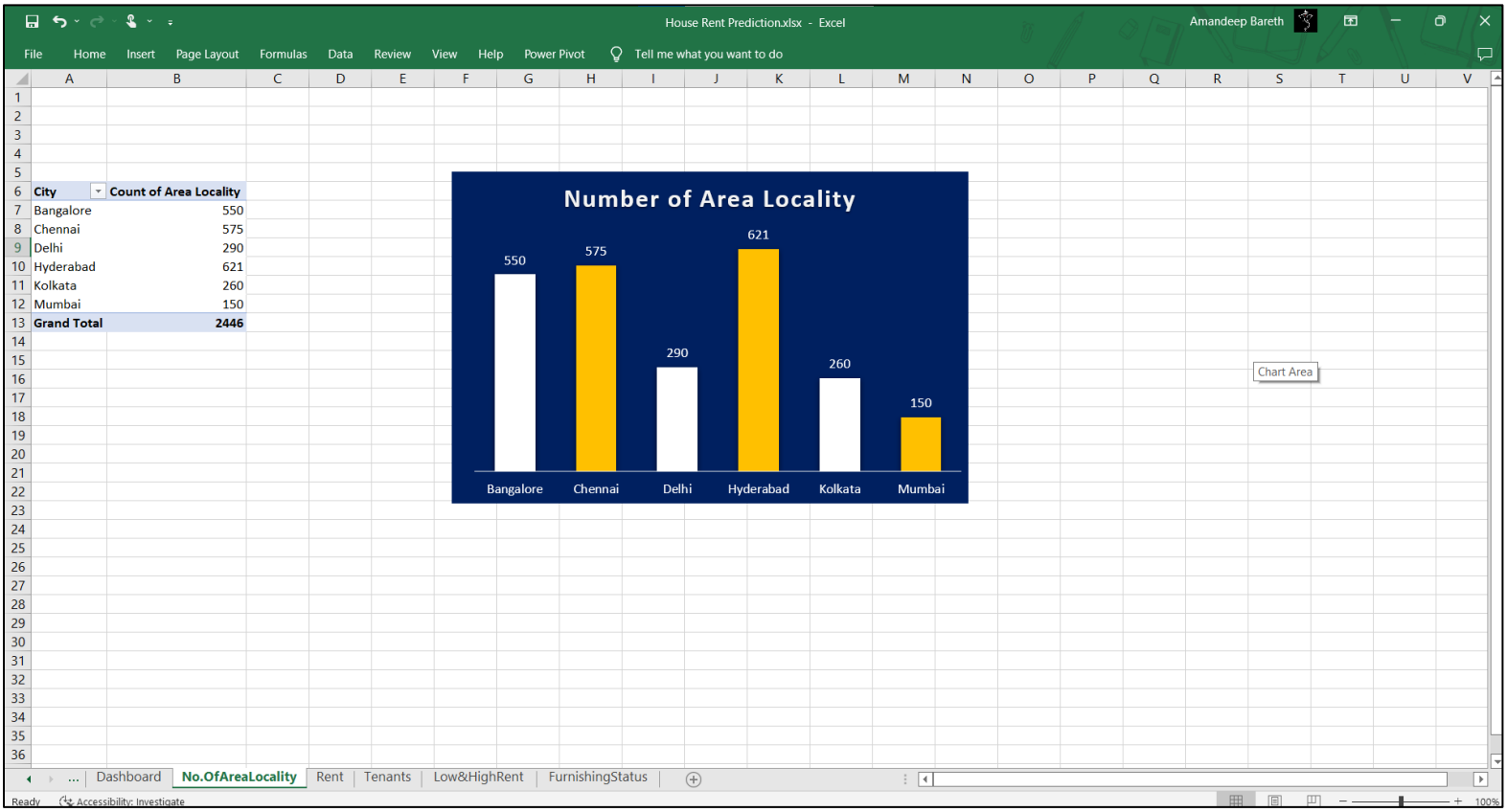
- Offers deep historical context for business.
- Enhances Business Intelligence solutions for decision making.
- Enables context and data aggregations so that business can generate higher revenue and/or save money.
- Enables a common data repository.
- Allows verification of data transformation, aggregation, and calculations rules.
- Allows sample data comparison between source and target system.
- Helps to improve productivity as it codifies and reuses without additional technical skills.

SCREENSHOT

1. Dashboard

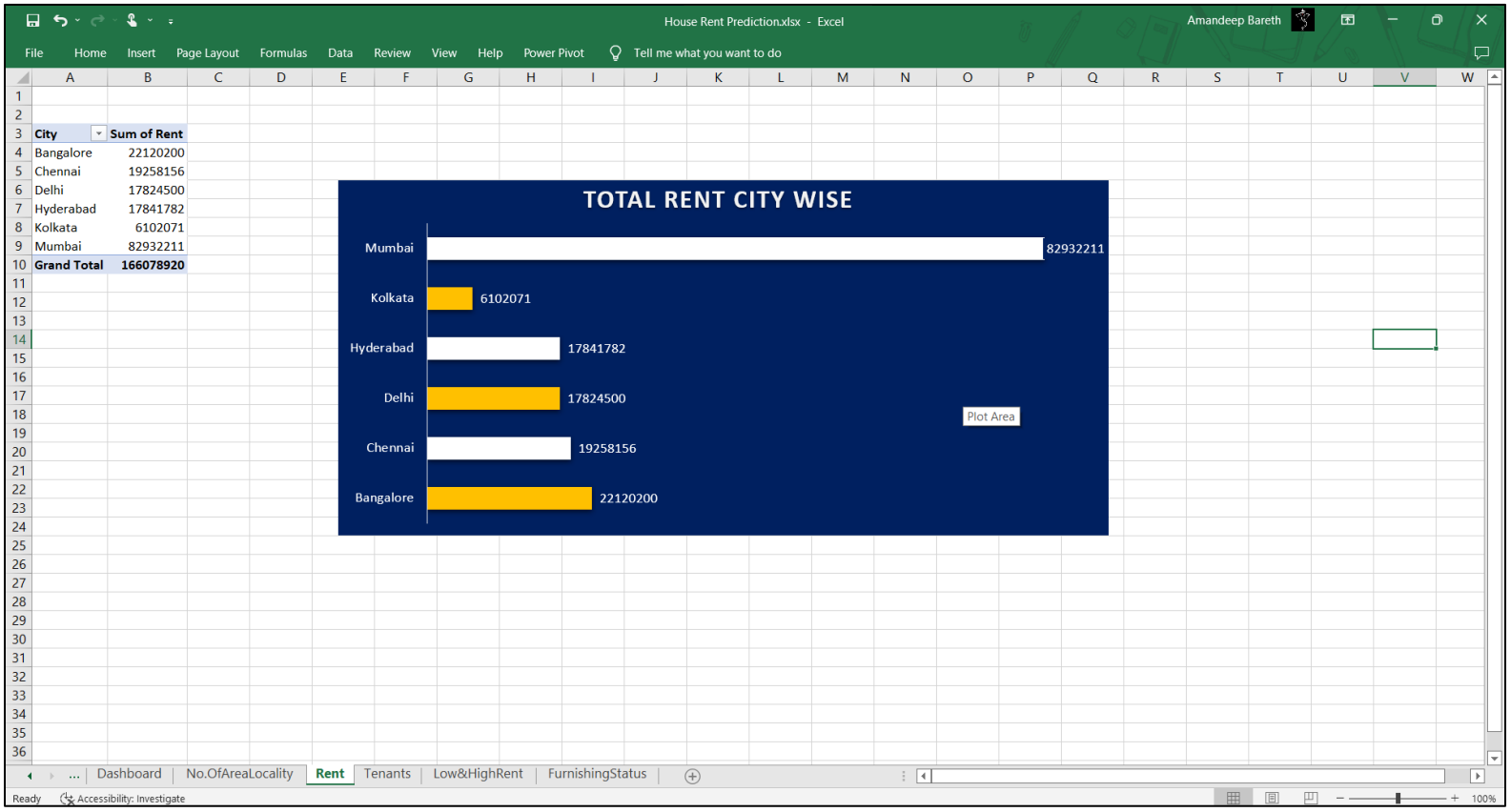


2. Number of Area Locality in Specific City



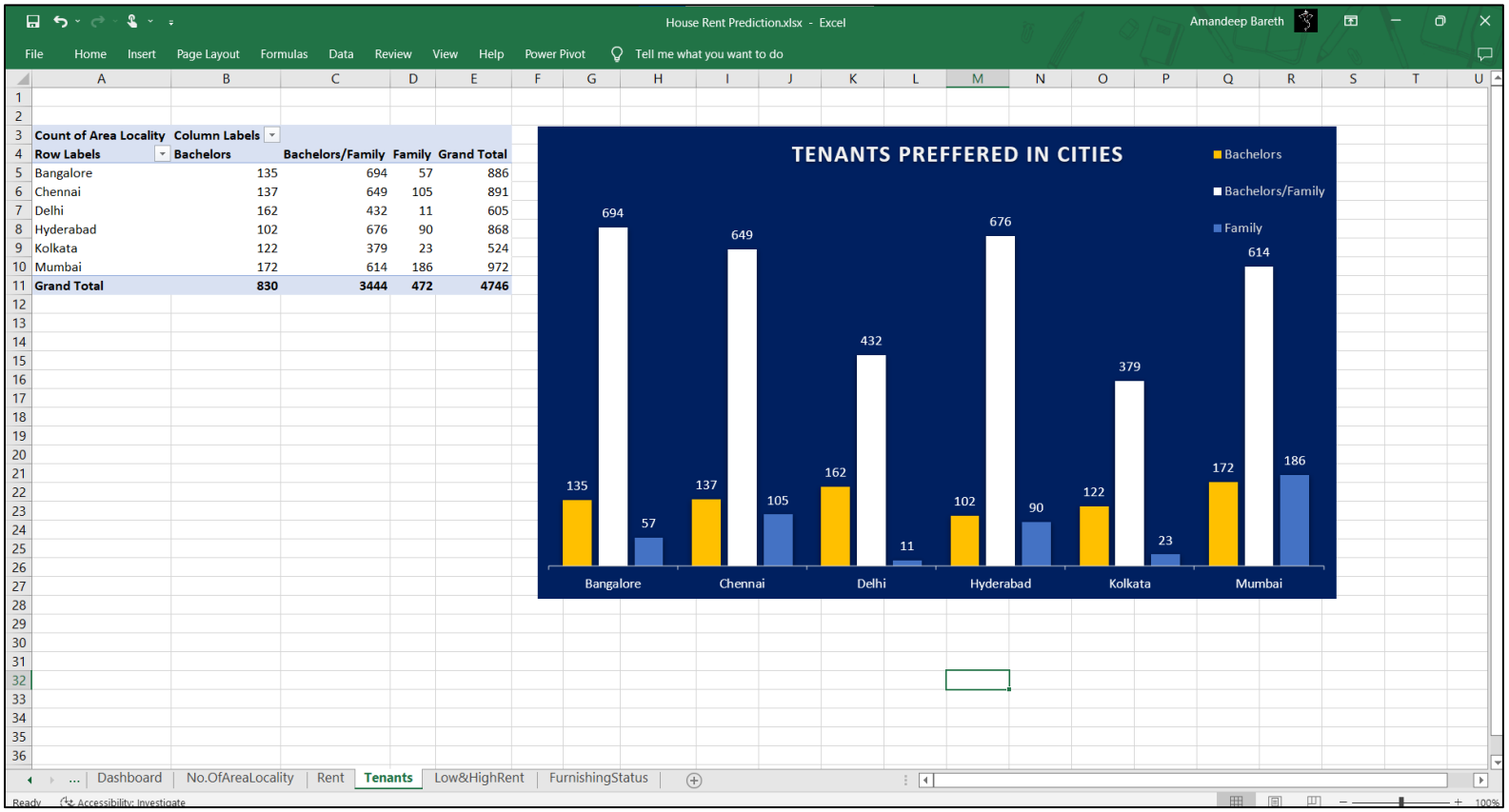
City	Count of Area Locality
Bangalore	550
Chennai	575
Delhi	290
Hyderabad	621
Kolkata	260
Mumbai	150
Grand Total	**2446**

3. Total Rent City Wise



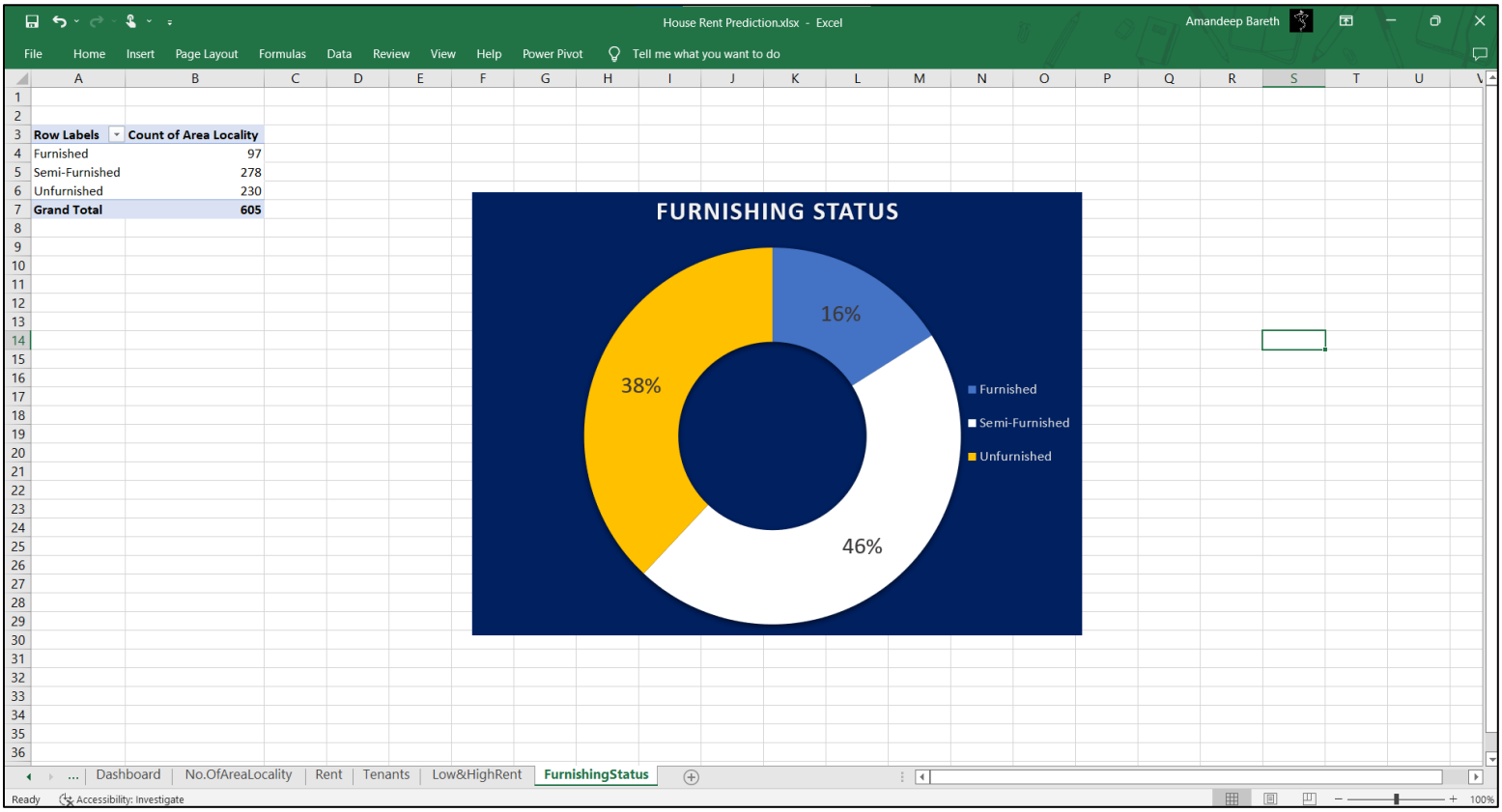
City	Sum of Rent
Bangalore	22120200
Chennai	19258156
Delhi	17824500
Hyderabad	17841782
Kolkata	6102071
Mumbai	82932211
Grand Total	**166078920**

4. Tenants Preferred in Specific City

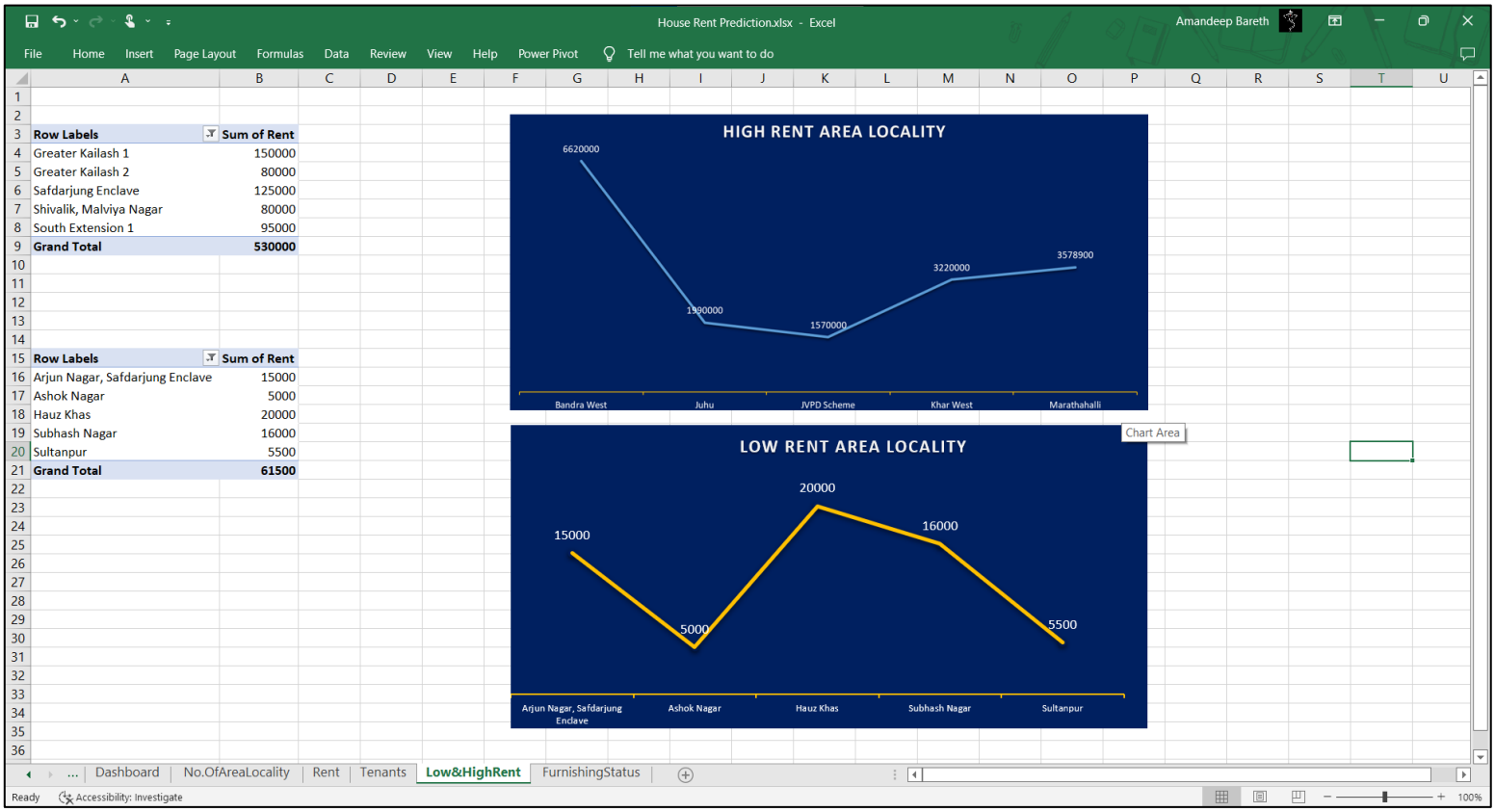


Count of Area Locality	Column Labels			
Row Labels	Bachelors	Bachelors/Family	Family	Grand Total
Bangalore	135	694	57	886
Chennai	137	649	105	891
Delhi	162	432	11	605
Hyderabad	102	676	90	868
Kolkata	122	379	23	524
Mumbai	172	614	186	972
Grand Total	**830**	**3444**	**472**	**4746**

5. Furnishing Status in Specific City



6. High Rent and Low Rent Area Locality City in Specific City



Row Labels	Sum of Rent
Greater Kailash 1	150000
Greater Kailash 2	80000
Safdarjung Enclave	125000
Shivalik, Malviya Nagar	80000
South Extension 1	95000
Grand Total	**530000**

Row Labels	Sum of Rent
Arjun Nagar, Safdarjung Enclave	15000
Ashok Nagar	5000
Hauz Khas	20000
Subhash Nagar	16000
Sultanpur	5500
Grand Total	**61500**

BIBLIOGRAPHY

1. To get idea about the Kaggle and its actual use

   https://www.quora.com/What-is-Kaggle-and-how-exactly-should-I-use-it

2. Kaggle dataset of house rent prediction

   https://www.kaggle.com/datasets/iamsouravbanerjee/house-rent-prediction-dataset

3. Trend of Rent

   https://www.magicbricks.com/

4. Talend Real-Time Open-Source Data Integration Software. 2020. What Is Extract, Transform, Load? Definition, Process and Tools

   https://www.talend.com/resources/what-is-etl/

5. How to create a dashboard in Excel

   https://clickup.com/blog/how-to-create-a-dashboard-in-excel/