INTRODUCTION TO DATA MANAGEMENT PROJECT REPORT

(Project Semester August-December 2022)

Flipkart Mobile Sales Analysis

Submitted by

Noothimadugu Sree Nimai

Registration No:12006507

Programme and Section: CSE KM100

Course Code: INT217

Under the Guidance of

Veerpal kaur:25909

Discipline of CSE/IT

Lovely School of Computer Science and Engineering

Lovely Professional University, Phagwara

CERTIFICATE

This is to certify that Noothimadugu Sree Nimai bearing Registration no. 12006507 has completed INT217 project titled, Flipkart mobile data analysis under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort, and study.

Signature and Name of the Supervisor
Designation of the Supervisor
School of Computer Science and Engineering
Lovely Professional University
Phagwara, Punjab.

Date: 6th November,2022

DECLARATION

I, Noothimadugu Sree Nimai, student of B.Tech 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: 6/11/2022

Signature

Registration No. 12006507

Noothimadugu Sree Nimai

ACKNOWLEDGEMENT:

I hereby express my heartiest thanks to all the sources who have contributed to the making of this project. I oblige thanks to all those who have supported, provided their valuable guidance, and helped for the accomplishment of this project. I also extend my hearty thanks to my family, friends, college teachers, and all the well-wishers.

I also would like to thank my project guide Veerpal kaur:25909 for her guidance and timely Suggestion and the information provided by her on this topic.

It is a matter of utmost pleasure to express my indebt and deep sense of gratitude to various people who extended their maximum help to supply the necessary information for the present thesis, which became available on account of the most selfless cooperation.

Above all its sincere thanks to the LOVELY PROFESSIONAL UNIVERSITY for which this project is given consideration and was done with utmost seriousness.

TABLE OF CONTENT

INTRODUCTION

About Flipkart:

Flipkart is an Indian e-commerce company, headquartered in Bangalore, Karnataka, India. It is the largest e-commerce company in India and was founded by Sachin and Binny Bansal. The company has a wide variety of products electronics like laptops, tablets, smartphones, and mobile accessories to in-vogue fashion staples like shoes, clothing, and lifestyle accessories; from modern furniture like sofa sets, dining tables, and wardrobes to appliances that make your life easy like washing machines, TVs, ACs, mixer grinder juicers and other time-saving kitchen and small appliances; from home furnishings like cushion covers, mattresses, and bedsheets to toys and musical instruments.

Mobile Phones

Mobile phones are one of the most rapidly rising industries, as well as one of the most prominent industries in the technology sector. The rate of increase has been exponential, with the number of mobile phone customers increasing fivefold in the last decade. Globally, the number of smartphones sold to end users climbed from 300 million in 2010 to 1.5 billion by 2020.

Flipkart and Mobile Phones

As previously stated, mobile phones are in high demand and are one of the ideal products for a novice to selling. Flipkart will be the ideal spot for a vendor to market their stuff because of its reach.

The dataset includes data on mobile phones from the top five most popular brands in India: Apple, Poco, Realme, Samsung, and Xiaomi. Information like RAM, ROM, Display Size. etc are present which distinguishes one product from another. At least one attribute distinguishes each product. Dataset has no null value.

Columns: There are 17 columns each having a title which is self-explanatory.

Rows: There are 430 rows each having a mobile with at least a distinct feature.

DESCRIPTION OF ATTRIBUTES:

Brand:

Brand name of the mobile

Model:

The model's name of the mobile.

base_color:

Color of mobile.

Processor:

Processor of mobile

screen_size:

screen sizes of mobiles are divided into

1. small
2. Medium
3. Large

ROM:

Read-only memory varying from 16GB to 256GB

RAM:

Random-access memory varying from 2 GB to 6GB

Display_size:

Size of the display of mobile in inches

Num_rear_camera

Number of back cameras

Num_front_camera:

Number of front cameras

Battery capacity:

The capacity of the battery in mAH.

Ratings:

Customer satisfaction rating on their overall mobile user experience (On a scale of 1 to 5).

Num_of_ratings:

Number of ratings for that particular mobile mobile

Sales_price:

The price of the mobile is ₹

Discount_percent:

Percentage of discount.

Sales:

Sales of that particular mobile.

Category:

Mobiles are divided into different categories according to their price ranges.

Objectives/Scope of the Analysis

- Best mobile according to price range
- Top-selling brands according to sales
- Rating of mobiles from top to bottom
- Mostly sold model sales according to brand
- Color-wise product count
- Brand-wise model count
- Sales according to the processor
- Sales according to screen size

SOURCE OF DATASET:

I collected this dataset for my project from the Kaggle website.

Kaggle:

Kaggle is an online community of data scientists and machine learners, owned by Google, 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 engineers, and enter competitions to solve data science challenges. Kaggle got its start by offering machine learning competitions and now also offers a public data platform, a cloud-based workbench for data science, and short-form AI education, on 8 March 2017, Google announced that it was acquiring Kaggle. The competition host prepares the data and a description of the problem. Participants experiment with different techniques and compete against each other to produce the best models. Work is shared publicly through Kaggle kernels to achieve a better benchmark and to inspire new ideas. Submissions can be made through Kaggle kernels, through manual upload, or by using the Kaggle API. For most competitions, submissions are scored immediately (based on their predictive accuracy relative to the hidden solution file) and summarized on a live leaderboard.

Reference link of dataset:

https://www.kaggle.com/datasets/shubhambathwal/flipkart-mobile-dataset

ETL Process:

ETL is a process in Data Warehousing, and it stands for Extract, Transform and Load. It is a process in which an ETL tool extracts the data from various data source systems, transforms it in the staging area, and then finally, loads it into the Data Warehouse system.

1) Data Extraction:

The first step of the ETL process is extraction. In this step, I removed all the duplicated values, false values, etc.,

2) Data Transformations:

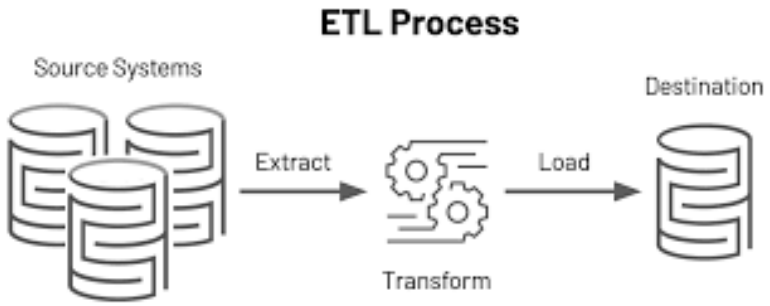
In this step, I replaced null values with NAN and transformed percentage values to numerical values, calculating some row values like categories that were not given from the dataset using their price ranges.

3) Data Load:

In this step, I transformed data according to my project and I had calculated

the required data and created pivot tables after data load.'

The process of ETL plays a key role in data integration strategies. ETL allows businesses to gather data from multiple sources and consolidate it into a single, centralized location. ETL also makes it possible for different types of data to work together.

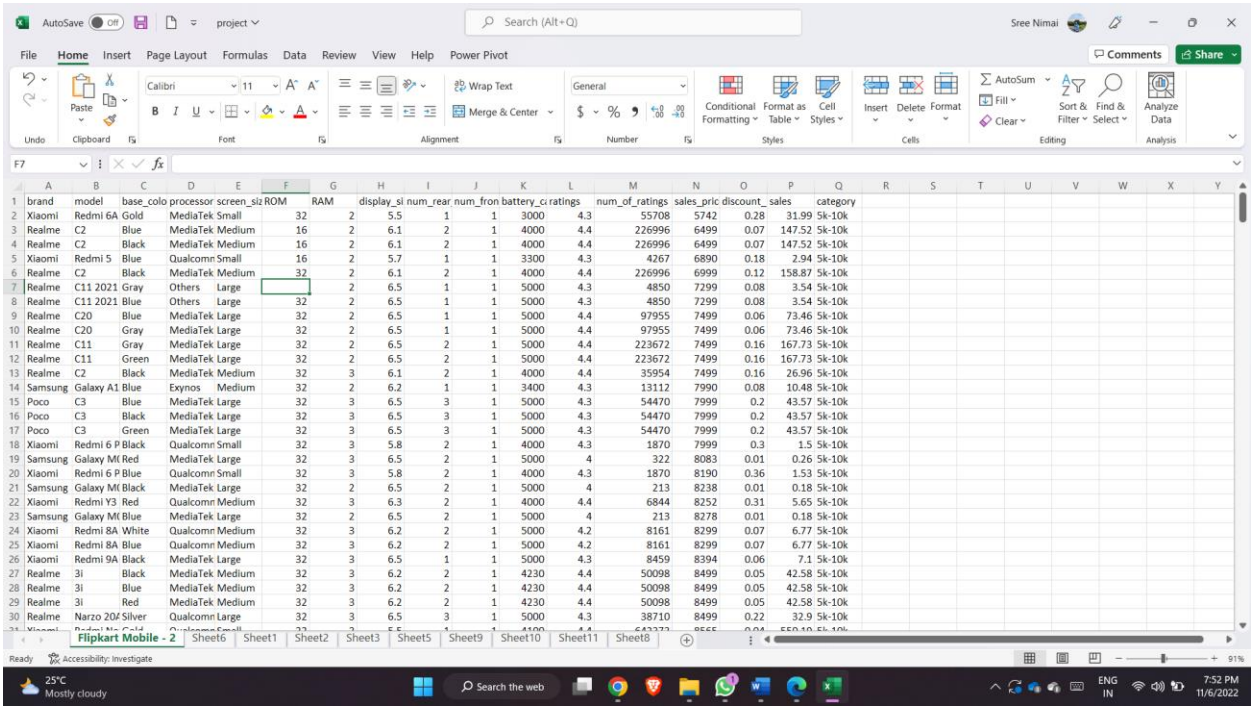


A typical ETL process collects and refines different types of data, then delivers the data to a data Warehouse such as Redshift, Azure, or Big Query.

ETL also makes it possible to migrate data between a variety of sources, destinations, and analysis tools. As a result, the ETL process plays a critical role in producing business intelligence and executing broader data management strategies.

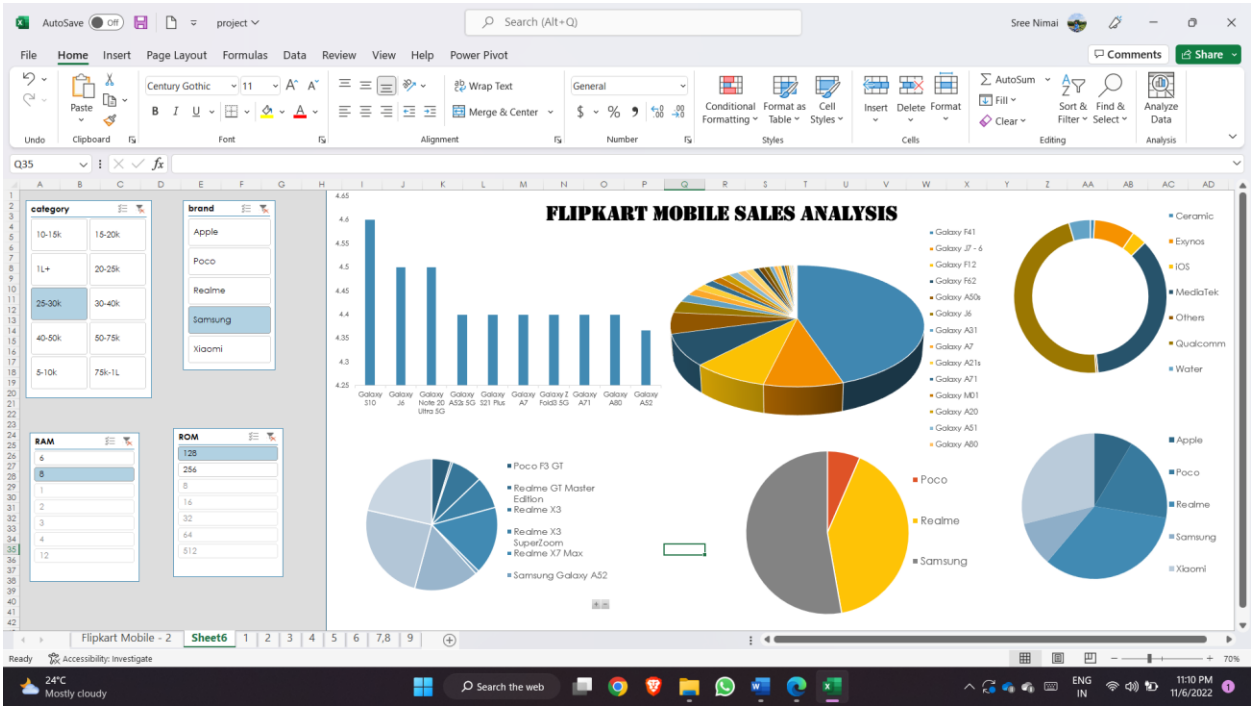
Three steps make up the ETL process and enable data to be integrated from source to destination. These are data extraction, data transformation, and data loading

Raw Data Set:



brand	model	base_colo	processor	screen_siz	ROM	RAM	display_si	num_rear	num_fron	battery_c	ratings	num_of_ratings	sales_pric	discount_	sales	category
Xiaomi	Redmi 6A	Gold	MediaTek	Small	32	2	5.5	1	1	3000	4.3	55708	5742	0.28	31.99	5k-10k
Realme	C2	Blue	MediaTek	Medium	16	2	6.1	2	1	4000	4.4	226996	6499	0.07	147.52	5k-10k
Realme	C2	Black	MediaTek	Medium	16	2	6.1	2	1	4000	4.4	226996	6499	0.07	147.52	5k-10k
Xiaomi	Redmi 5	Blue	Qualcomn	Small	16	2	5.7	1	1	3300	4.3	4267	6890	0.18	2.94	5k-10k
Realme	C2	Black	MediaTek	Medium	32	2	6.1	2	1	4000	4.4	226996	6999	0.12	158.87	5k-10k
Realme	C11 2021	Gray	Others	Large		2	6.5	1	1	5000	4.3	4850	7299	0.08	3.54	5k-10k
Realme	C11 2021	Blue	Others	Large	32	2	6.5	1	1	5000	4.3	4850	7299	0.08	3.54	5k-10k
Realme	C20	Blue	MediaTek	Large	32	2	6.5	1	1	5000	4.4	97955	7499	0.06	73.46	5k-10k
Realme	C20	Gray	MediaTek	Large	32	2	6.5	1	1	5000	4.4	97955	7499	0.06	73.46	5k-10k
Realme	C11	Gray	MediaTek	Large	32	2	6.5	2	1	5000	4.4	223672	7499	0.16	167.73	5k-10k
Realme	C11	Green	MediaTek	Large	32	2	6.5	2	1	5000	4.4	223672	7499	0.16	167.73	5k-10k
Realme	C2	Black	MediaTek	Medium	32	3	6.1	2	1	4000	4.4	35954	7499	0.16	26.96	5k-10k
Samsung	Galaxy A1	Blue	Exynos	Medium	32	2	6.2	1	1	3400	4.3	13112	7990	0.08	10.48	5k-10k
Poco	C3	Blue	MediaTek	Large	32	3	6.5	3	1	5000	4.3	54470	7999	0.2	43.57	5k-10k
Poco	C3	Black	MediaTek	Large	32	3	6.5	3	1	5000	4.3	54470	7999	0.2	43.57	5k-10k
Poco	C3	Green	MediaTek	Large	32	3	6.5	3	1	5000	4.3	54470	7999	0.2	43.57	5k-10k
Xiaomi	Redmi 6 P	Black	Qualcomn	Small	32	3	5.8	2	1	4000	4.3	1870	7999	0.3	1.5	5k-10k
Samsung	Galaxy M(	Red	MediaTek	Large	32	3	6.5	2	1	5000	4	322	8083	0.01	0.26	5k-10k
Xiaomi	Redmi 6 P	Blue	Qualcomn	Small	32	3	5.8	2	1	4000	4.3	1870	8190	0.36	1.53	5k-10k
Samsung	Galaxy M(	Black	MediaTek	Large	32	2	6.5	2	1	5000	4	213	8238	0.01	0.18	5k-10k
Xiaomi	Redmi Y3	Red	Qualcomn	Medium	32	3	6.3	2	1	4000	4.4	6844	8252	0.31	5.65	5k-10k
Samsung	Galaxy M(	Blue	MediaTek	Large	32	2	6.5	2	1	5000	4	213	8278	0.01	0.18	5k-10k
Xiaomi	Redmi 8A	White	Qualcomn	Medium	32	3	6.2	2	1	5000	4.2	8161	8299	0.07	6.77	5k-10k
Xiaomi	Redmi 8A	Blue	Qualcomn	Medium	32	3	6.2	2	1	5000	4.2	8161	8299	0.07	6.77	5k-10k
Xiaomi	Redmi 9A	Black	MediaTek	Large	32	3	6.5	1	1	5000	4.3	8459	8394	0.06	7.1	5k-10k
Realme	3i	Black	MediaTek	Medium	32	3	6.2	2	1	4230	4.4	50098	8499	0.05	42.58	5k-10k
Realme	3i	Blue	MediaTek	Medium	32	3	6.2	2	1	4230	4.4	50098	8499	0.05	42.58	5k-10k
Realme	3i	Red	MediaTek	Medium	32	3	6.2	2	1	4230	4.4	50098	8499	0.05	42.58	5k-10k
Realme	Narzo 20A	Silver	Qualcomn	Large	32	3	6.5	3	1	5000	4.3	38710	8499	0.22	32.9	5k-10k

Dashboard:


FLIPKART MOBILE SALES ANALYSIS
category
10-15k
15-20k
1L+
20-25k
25-30k
30-40k
40-50k
50-75k
5-10k
75k-1L
brand
Apple
Poco
Realme
Samsung
Xiaomi
RAM
6
8
1
2
3
4
12
ROM
128
256
8
16
32
64
512
Poco F3 GT
Realme GT Master Edition
Realme X3
Realme X3 SuperZoom
Realme X7 Max
Samsung Galaxy A52
Poco
Realme
Samsung
Ceramic
Exynos
IOS
MediaTek
Others
Qualcomm
Water
Apple
Poco
Realme
Samsung
Xiaomi
Flipkart Mobile - 2
Sheet6

Analysis of dataset (1st Objective):

Objective:

This objective is to tell good-selling brands according to price ranges

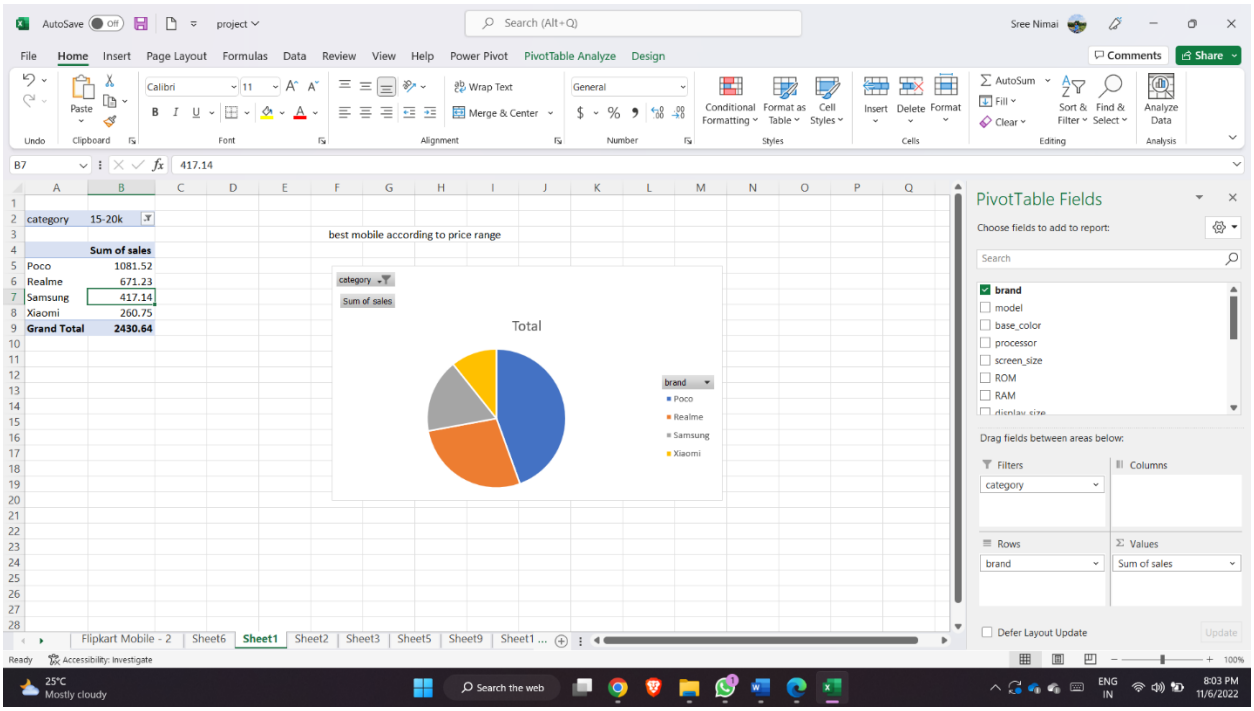
General Description:

I have used brands as rows and values as sales. In sales used the field setting as a sum to get the total count. And the last filter is a category so that mobiles can be divided according to their price ranges

Specific Requirements, functions, and formulas:

Brand, category, and the sum of sales are required to measure this pivot chart.

Analysis results and visualization: The pivot chart represents the mobile name as their price ranges.



Analysis of dataset (2nd Objective):

Objective:

This objective is to top-selling brands according to sales

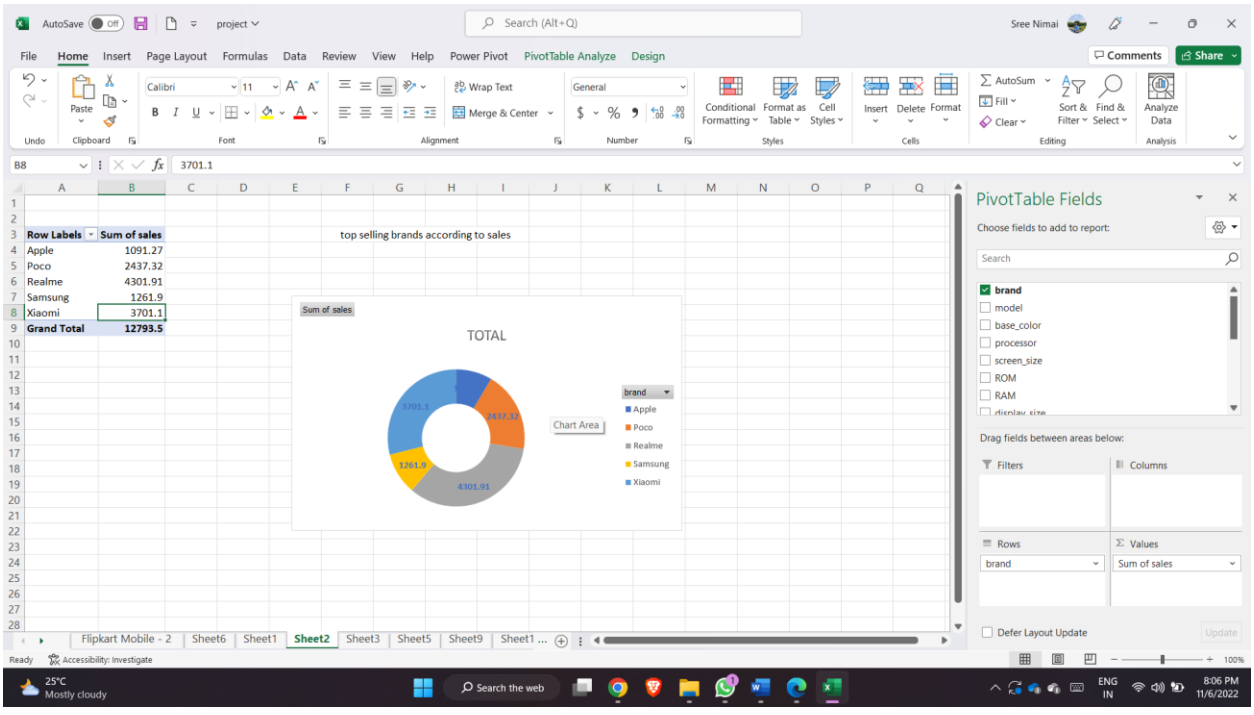
General Description:

I have used brands as rows and values as sales. In sales used the field setting as a sum to get the total count.

Specific Requirements, functions, and formulas:

Brand and sum of sales are required to measure this pivot chart.

Analysis results and visualization: The pivot chart represents the top-selling brands according to their sales.



Analysis of dataset (3rd Objective):

Objective:

This objective is to tell the rating of mobiles from top to bottom

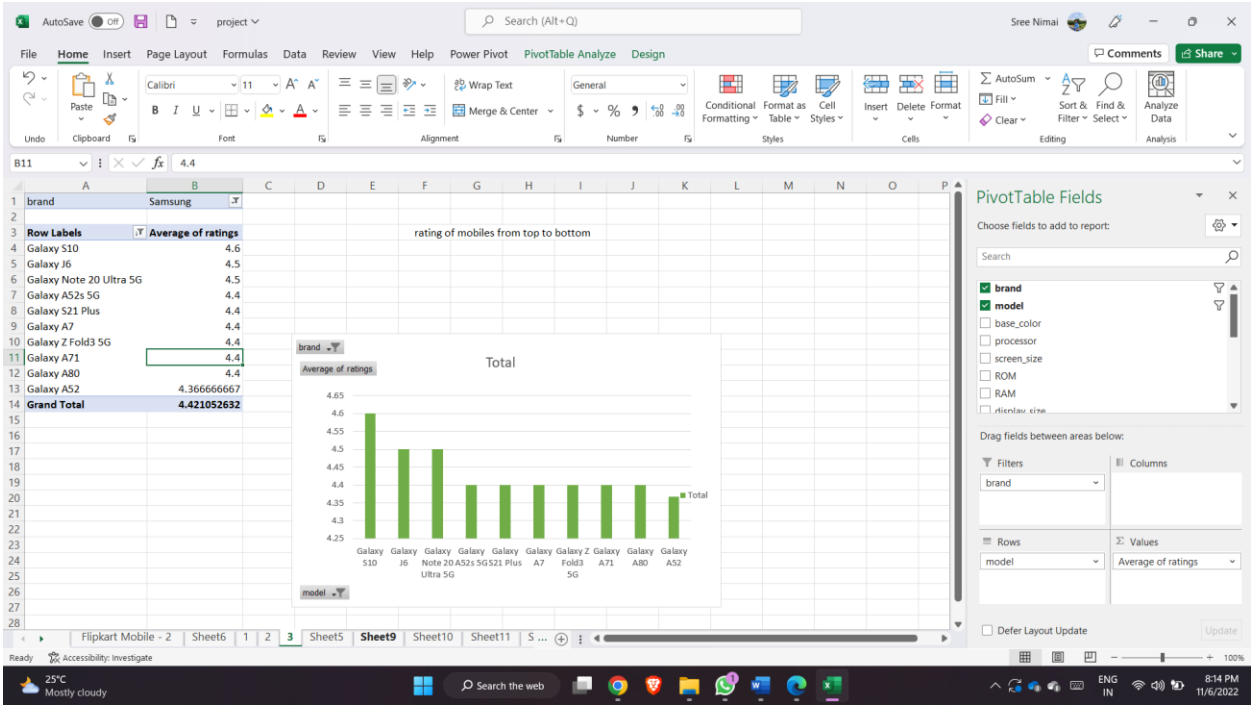
General Description:

I have used an average rating as a value field setting so that I can accurate results of the rating. And filter as a brand so divides them into their respective brands.

Specific Requirements, functions, and formulas:

Brand, model, and an average of ratings are required to measure this pivot chart.

Analysis results and visualization: The pivot chart represents a rating of models from top to bottom



Analysis of dataset (4th Objective):

Objective:

This objective is to tell the mostly sold model according to brand

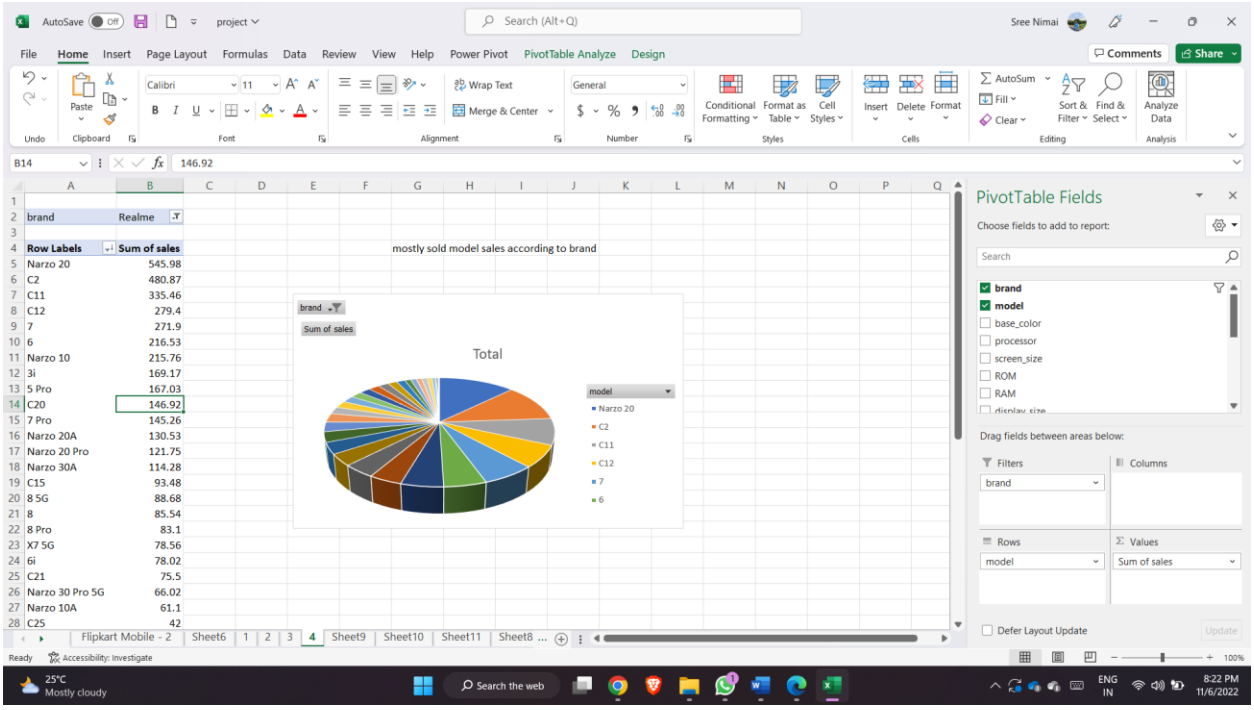
General Description:

I have used an average rating as a value field setting so that I can accurately results of the rating. And filter as a brand so divides them into their respective brands.

Specific Requirements, functions, and formulas:

Brand, model, and an average of ratings are required to measure this pivot chart.

Analysis results and visualization: The pivot chart represents the most sold models according to brand



Analysis of dataset (5th Objective):

Objective:

This objective is to tell the mostly sold model according to brand

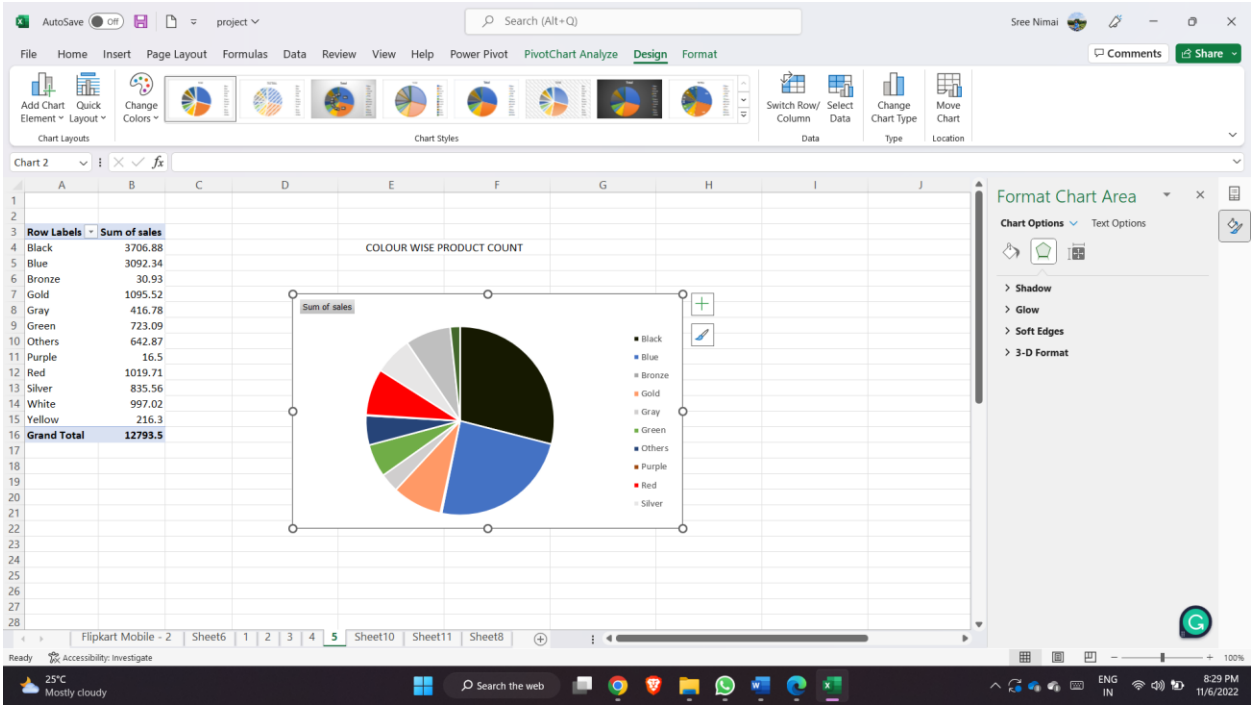
General Description:

I have used an average rating as a value field setting so that I can accurately results of the rating. And filter as a brand so divides them into their respective brands.

Specific Requirements, functions, and formulas:

Brand, model, and an average of ratings are required to measure this pivot chart.

Analysis results and visualization: The pivot chart represents the colorwise product count



Analysis of dataset (6th Objective):

Objective:

This objective is to tell the brand-wise model count

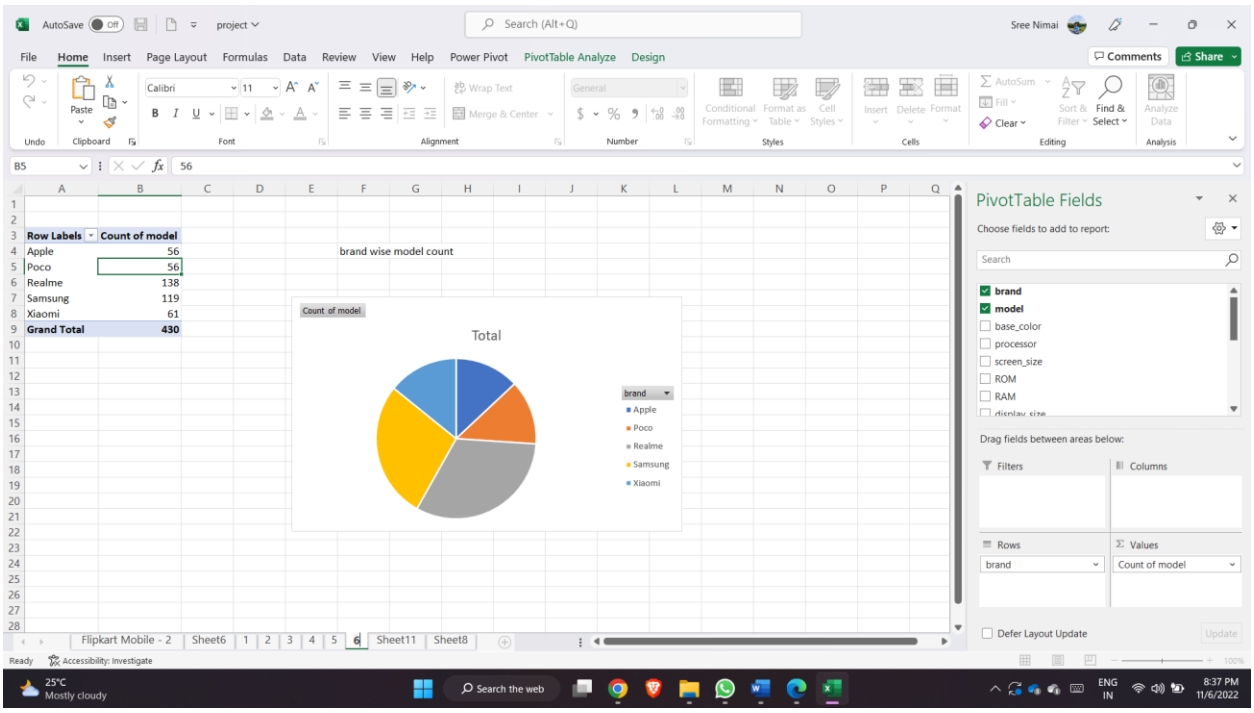
General Description:

I have used Brand as a row and the count of the model as values. I have used count to count the number of models according to brands.

Specific Requirements, functions, and formulas:

Brand and count of model ratings are required to measure this pivot chart.

Analysis results and visualization: The pivot chart represents the number of models according to brand



Analysis of dataset (7th Objective):

Objective:

This objective is to tell the sales according to the processor

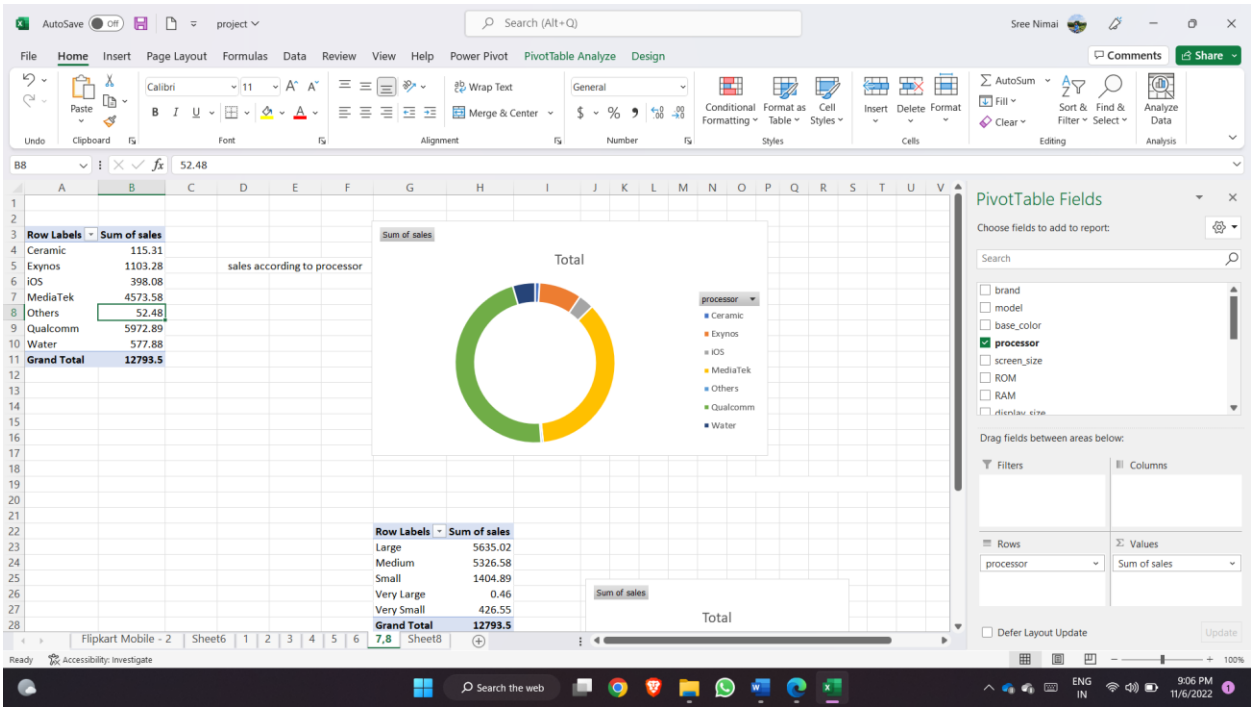
General Description:

I have used the processor as a row and the sum of sales as values. I used to sum to know the total sales of that processor.

Specific Requirements, functions, and formulas:

processor and sum of sales are required to measure this pivot chart.

Analysis results and visualization: The pivot chart represents the mostly selling processor



Analysis of dataset (8th Objective):

Objective:

This objective is to tell the sales according to the screen size

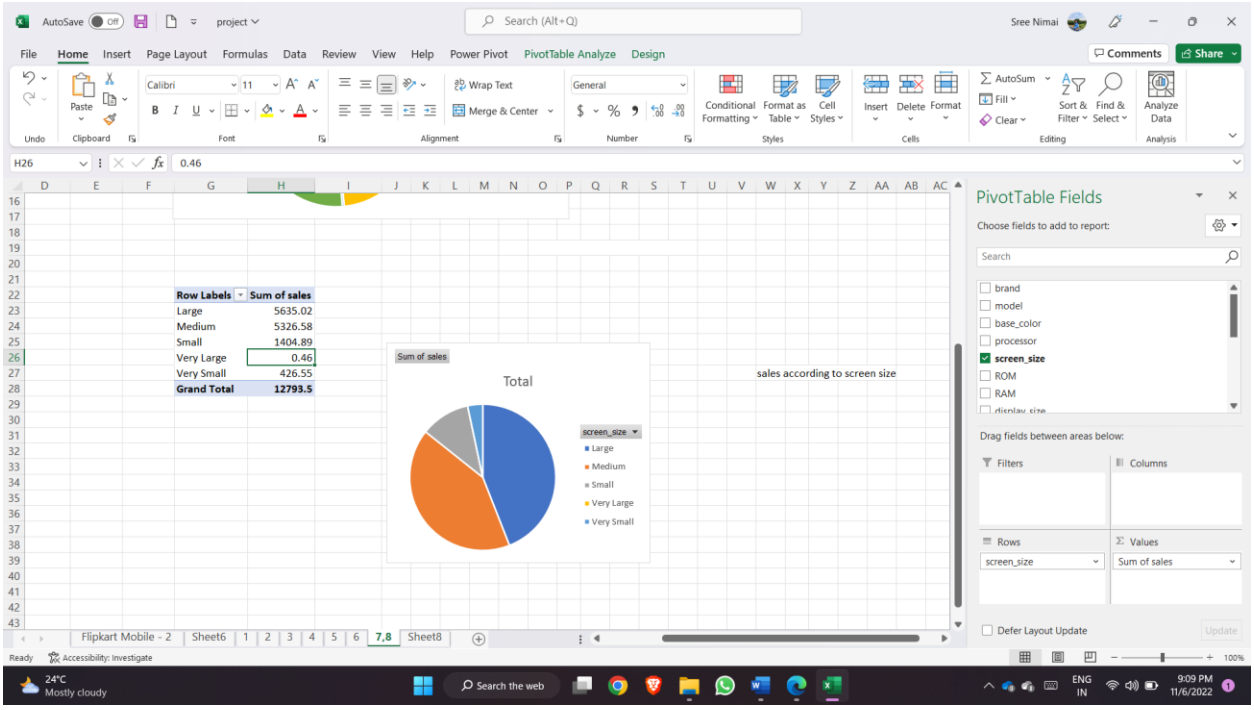
General Description:

I have used the screen size as a row and the sum of sales as values. I have used sum to know the total sales of that screen size.

Specific Requirements, functions, and formulas:

Screen size and the sum of sales are required to measure this pivot chart.

Analysis results and visualization: The pivot chart represents the most sales of mobile by screen sizes



Analysis of dataset (9th Objective):

Objective:

This objective is to sort according to user specifications.

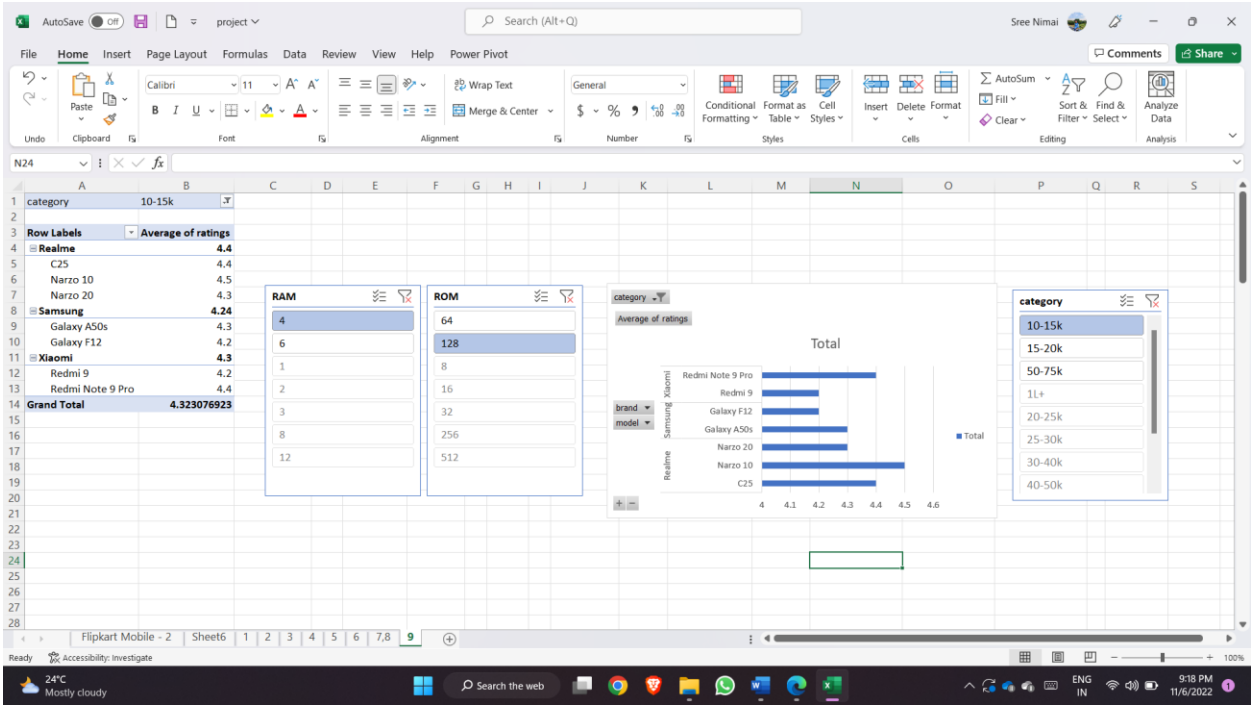
General Description:

I have used brand, and model as row, an average of rating as value, and category as a filter. By this user can be able to get a mobile name according to his specifications like RAM, ROM, and Cost.

Specific Requirements, functions, and formulas:

Brand, model, category, and average rating are required to measure this pivot chart.

Analysis results and visualization: The pivot chart represents the RAM, ROM, Cost wise mobile names.



List of Analysis with results:

1)Realme is having most sales followed by Xiaomi when compared to all other brands

2)Black is the most liked color for customers according to their sales.

3)Realme is the brand that is producing more models for the market.

4)According to results Qualcomm is the highest-selling processor.

5)When comparing display sizes Large and medium are having similar sales in the market.

Reference/ Bibliography:

- https://www.kaggle.com/datasets/shubhambathwal/flipkart-mobile-dataset
- https://youtu.be/5eLtjO2Hfs0