

**Project Problem Statement**

This analysis is aimed towards prospective as well as existing OTT platform providers who are planning an expansion in the field and need an overview about the current situation and demand of public so as to strategize their business methodology.

Over the Top (OTT) platform is the new age content provider which has revolutionized the way the common public views and interacts with entertainment. On demand and having a wide range of selection has drawn the public eye towards it. It is a budding industry worthy of investment, and we want to help the upcoming entrepreneurs in the field and the existing industrialists with all-over analysis of present mass preferences to better understand the industry requirements and demands.

**Methodology**

Our aim is to provide an in-depth analysis of people’s personal preferences revolving around OTT platforms, which involves everything; from their choice of genres to their preferred OTT platform, from their favorite movies to their preferred language to watch movies/shows in. Recording prospective customers’ personal information like location, gender, age, etc. is also of utmost importance for such analysis as personal information serves as the guiding light for businesses to plan ahead and scheme out their further projects to cater to the target audience, which will in turn increase company profits.

After collecting the desired data, we focus on finding any trends in data which might lead us to any certain conclusion. We use the various categories and find whether they are related and how. The effect of one category on another category is critical information to make presumptions about the trend of the data collected. These results are very useful to figure out the taste and choice of the respondents, who would ultimately be the target audience.

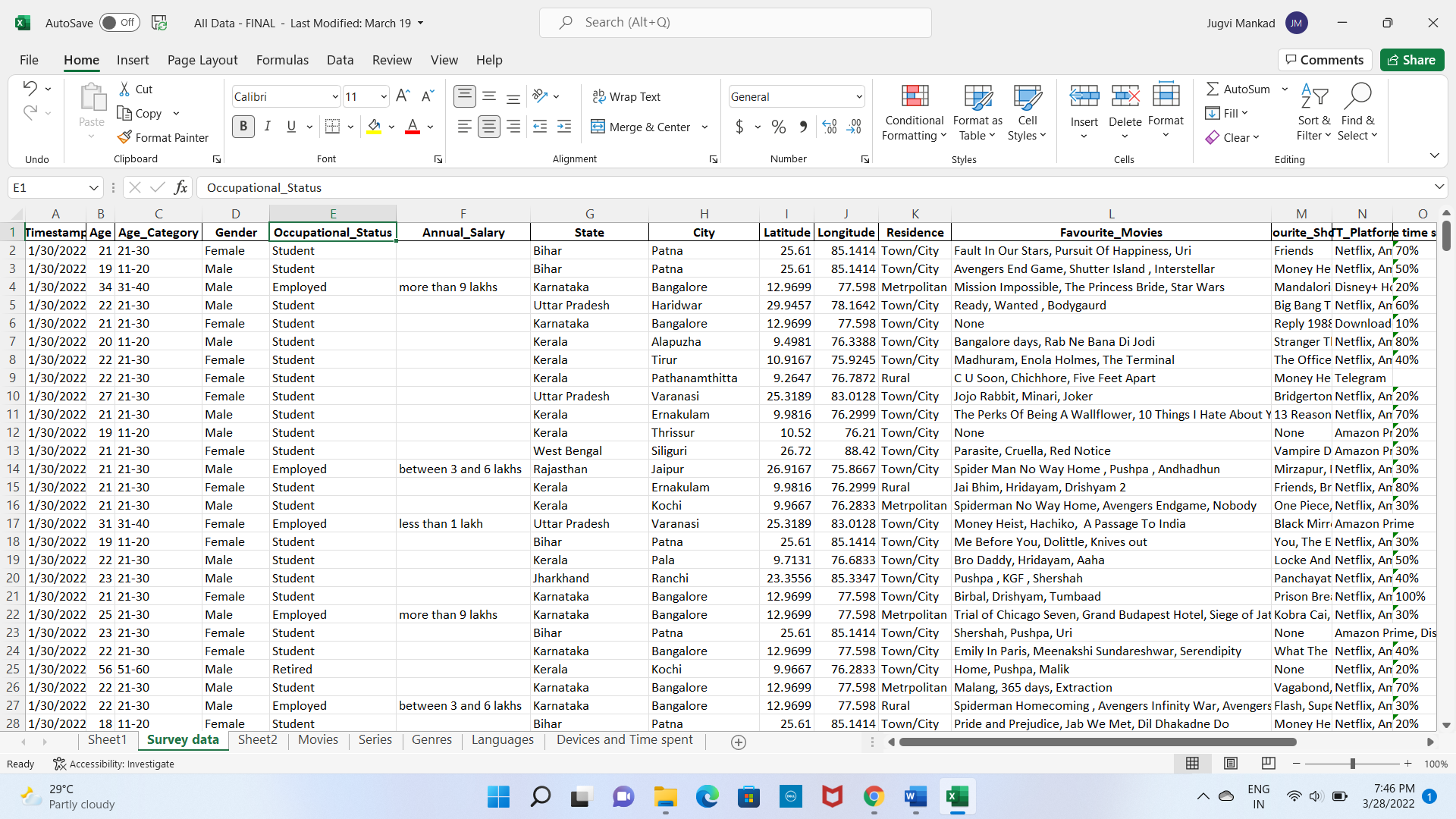
Next, we test the data to see whether our common presumptions are true or not, or how different fields are dependent on each other and to how much degree. Different tests are used to test different types of data. We also try to figure out if and how our data follows a certain distribution, which would enable us to figure out multiple possibilities and probabilities of various scenarios relating to our data.

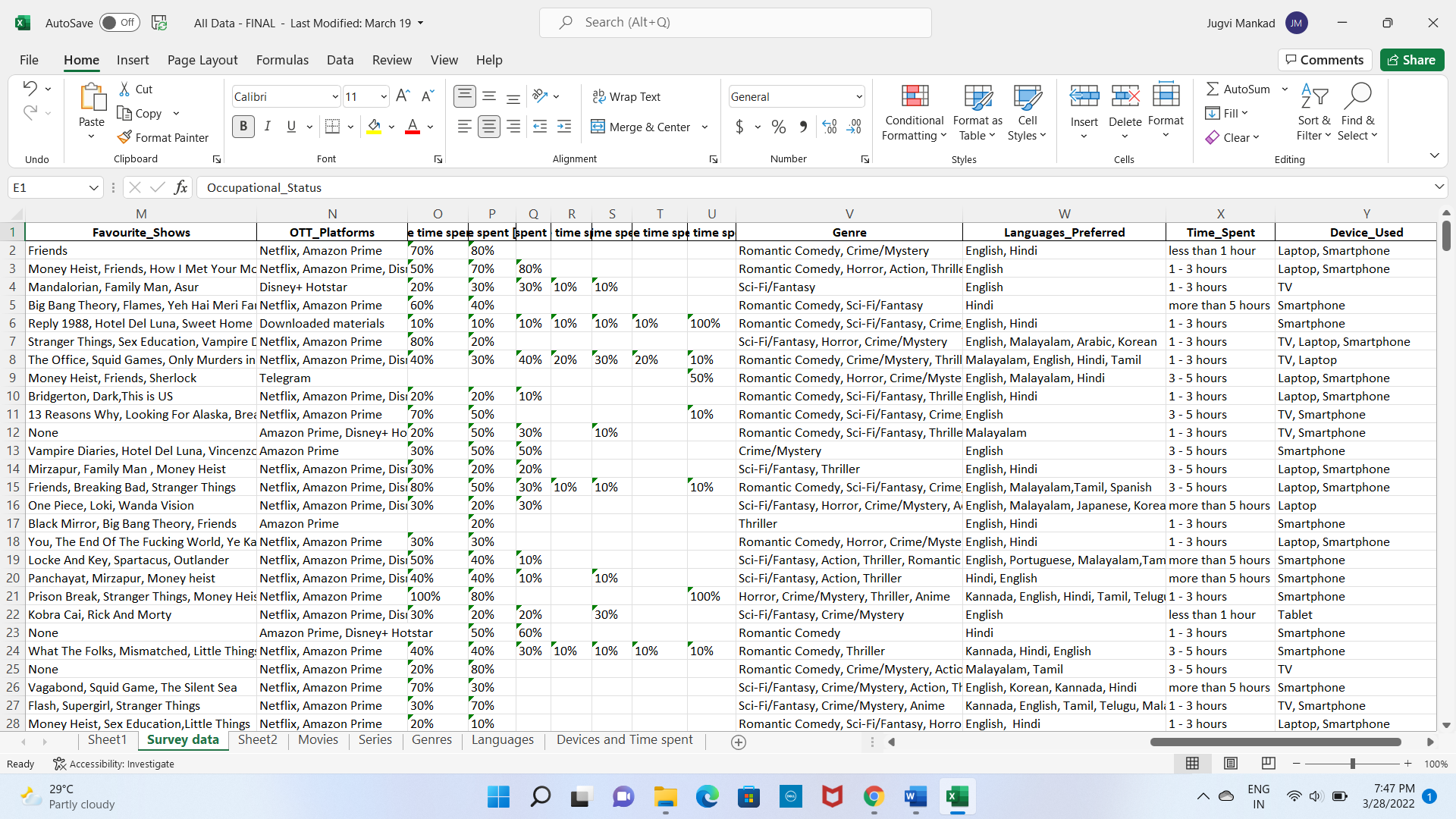
**Source of Datasets**

We used a primary data source to collect our data.

We prepared a Google Form with our required fields and recorded the responses.

The collected information is stored in a Microsoft Excel sheet. A picture of the raw data collected is attached below.





This raw data was then cleaned and filtered, and from it, other multiple datasets were created for ease of usage.

**Data Cleaning:**

* We used python pandas library to clean the data.
* We also used Microsoft Excel for filtering and creating pivot tables for ease of understanding.
* A field named ‘Percentage time spent on each platform’ was dropped as it didn’t get the expected responses, rendering it useless.
* A field named ‘Annual Salary’ was dropped as it received very low percentage of response, hence it couldn’t be used.

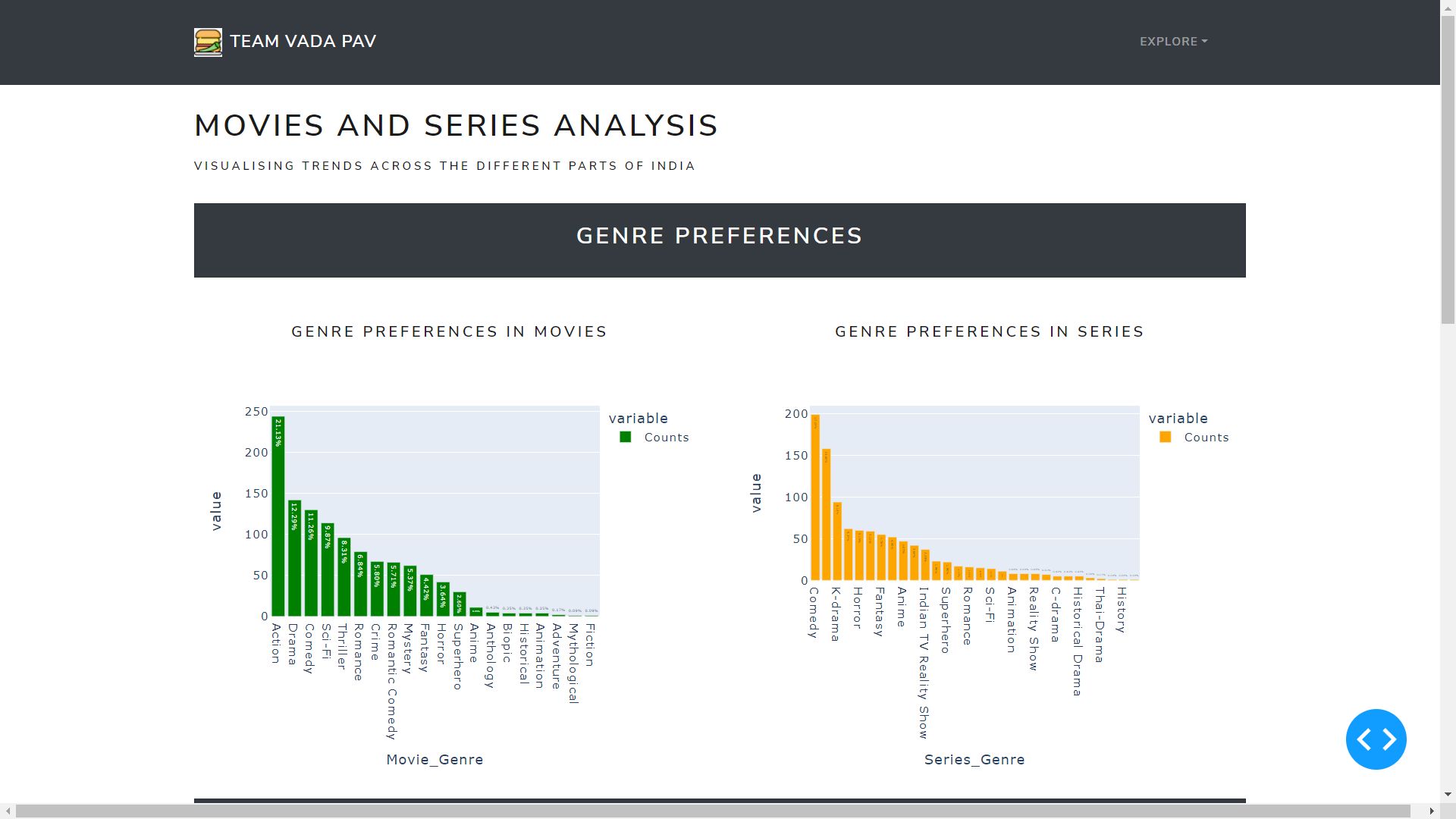
**Explanation:**

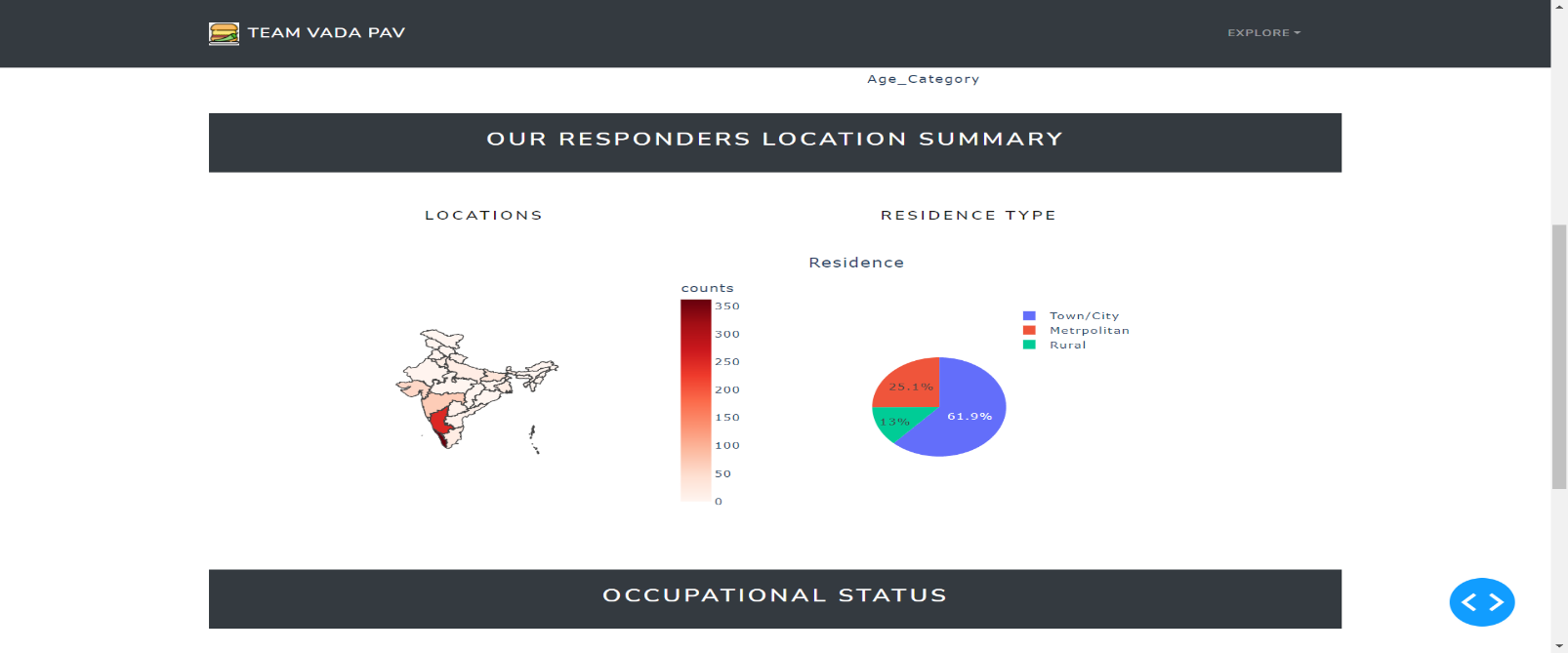
Our entire model depends on the fact that we have access to people’s personal information. This personal information is the most crucial component on which the entire approach depends.

All this information is used to figure out if and how one’s personal information affects their choices regarding OTT platforms and its related fields.

Using visual graphs:

The following graphs were plotted to study the relation between the variables:

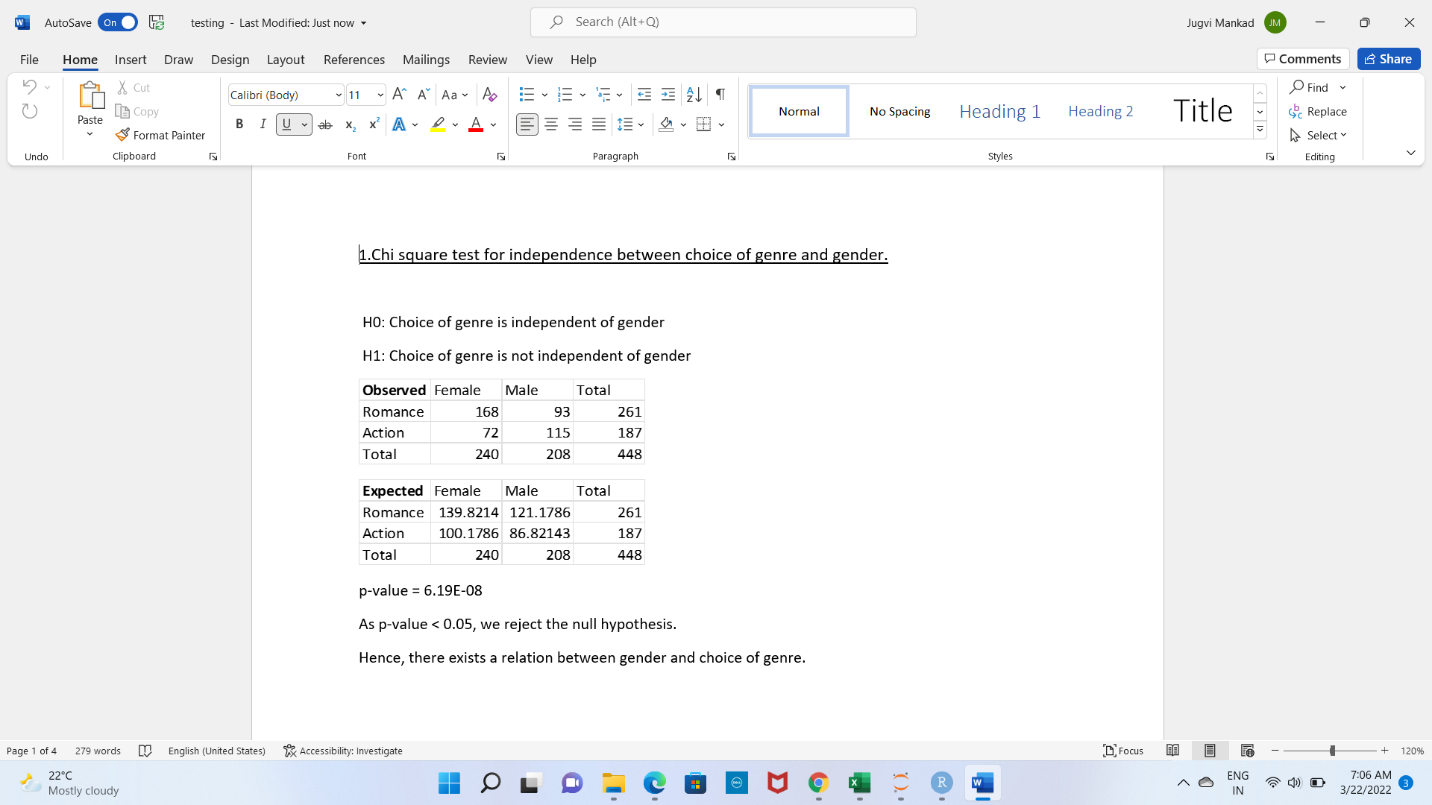
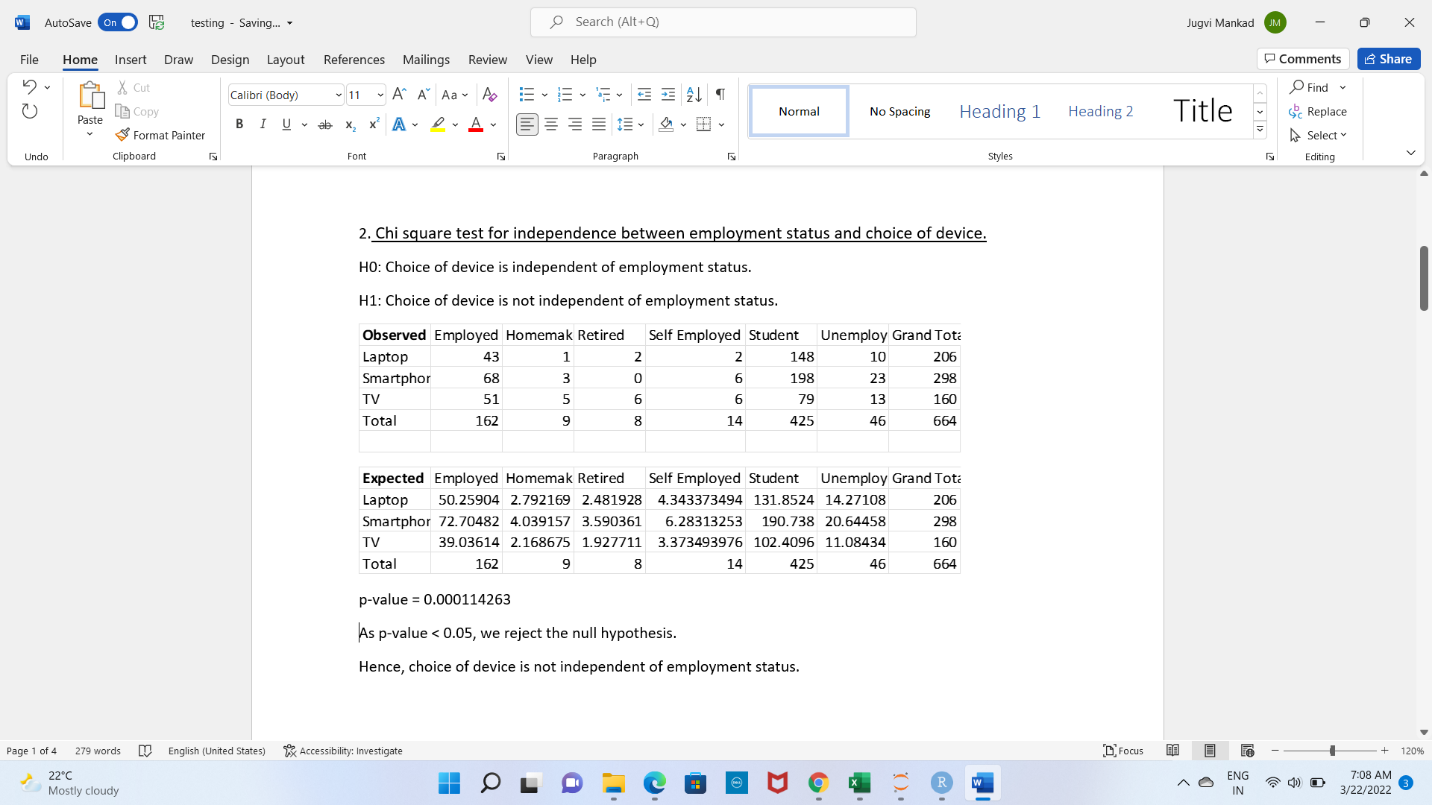




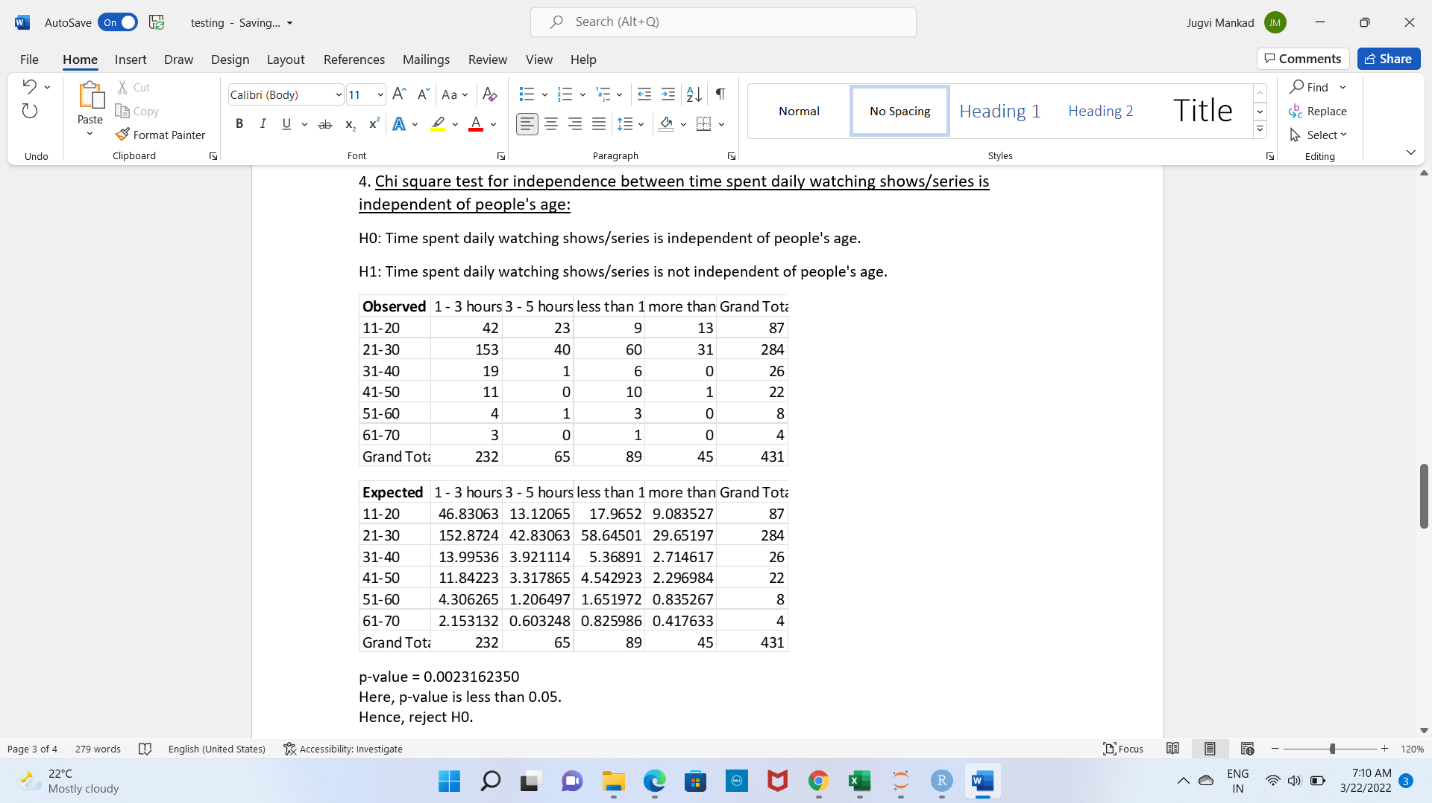
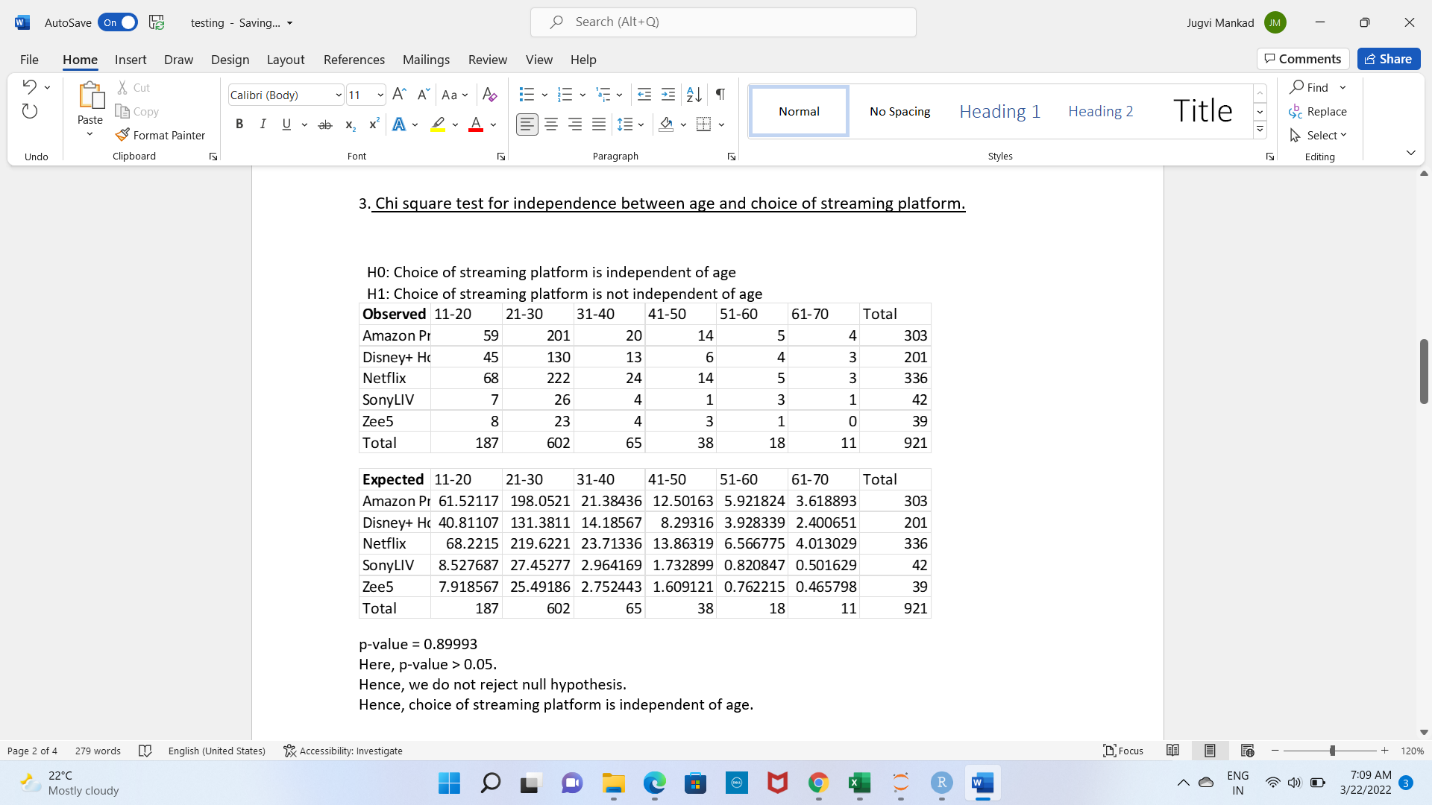
**Chi-Square Testing:**

We performed chi-square test for independence on various categorical variables to check whether they are dependent on each other or not.

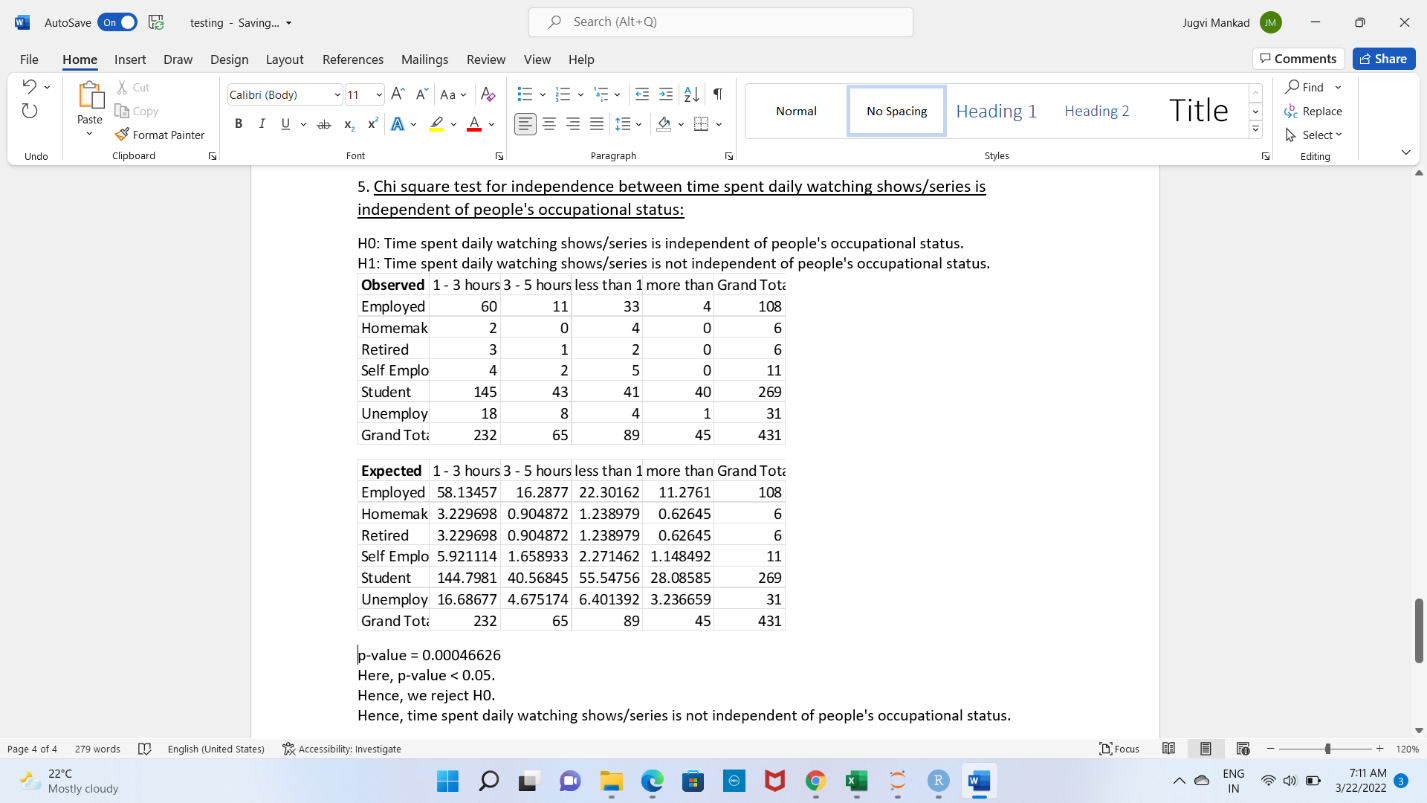
1. **‘Genre’ and ‘Gender’: 3. ‘Device’ and ‘Employment Status’:**

1. **‘Streaming platform’ and ‘Age’: 4. ‘Time spent’ and ‘Age’**



1. **‘Time spent on OTT platforms in a day’ and ‘Occupational Status’:**



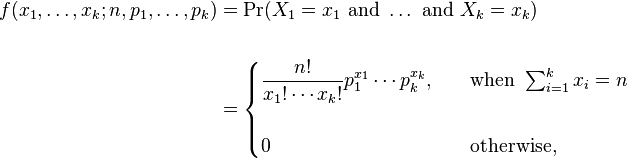
**Multinomial Distribution:**

We observed that the data we collected follows a multinomial distribution.

For each field, we have ‘n’ number of trials- which are the number of responses, and ‘k’ possible outcomes-which is the range of that particular field.

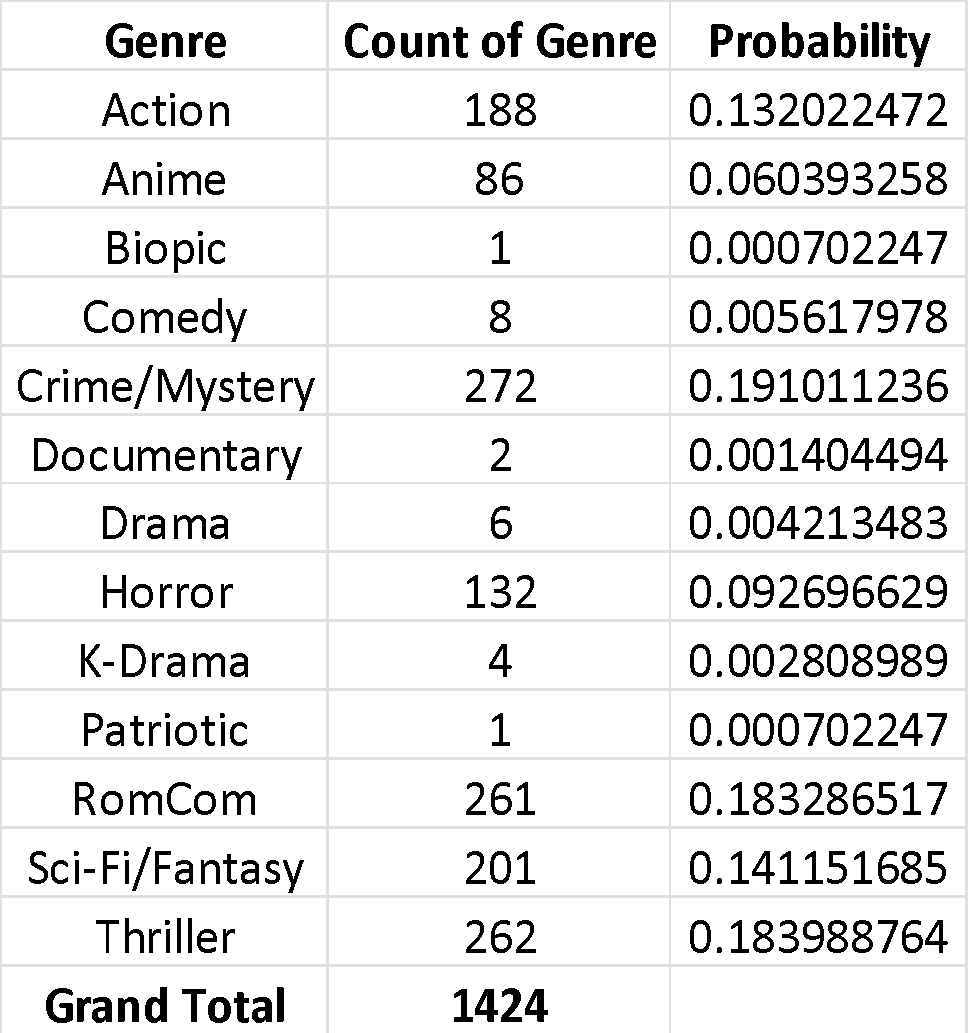
All ‘k’ outcomes are mutually exclusive and probability of each ‘k’ trial remains the same.

The formula to calculate the probability is:

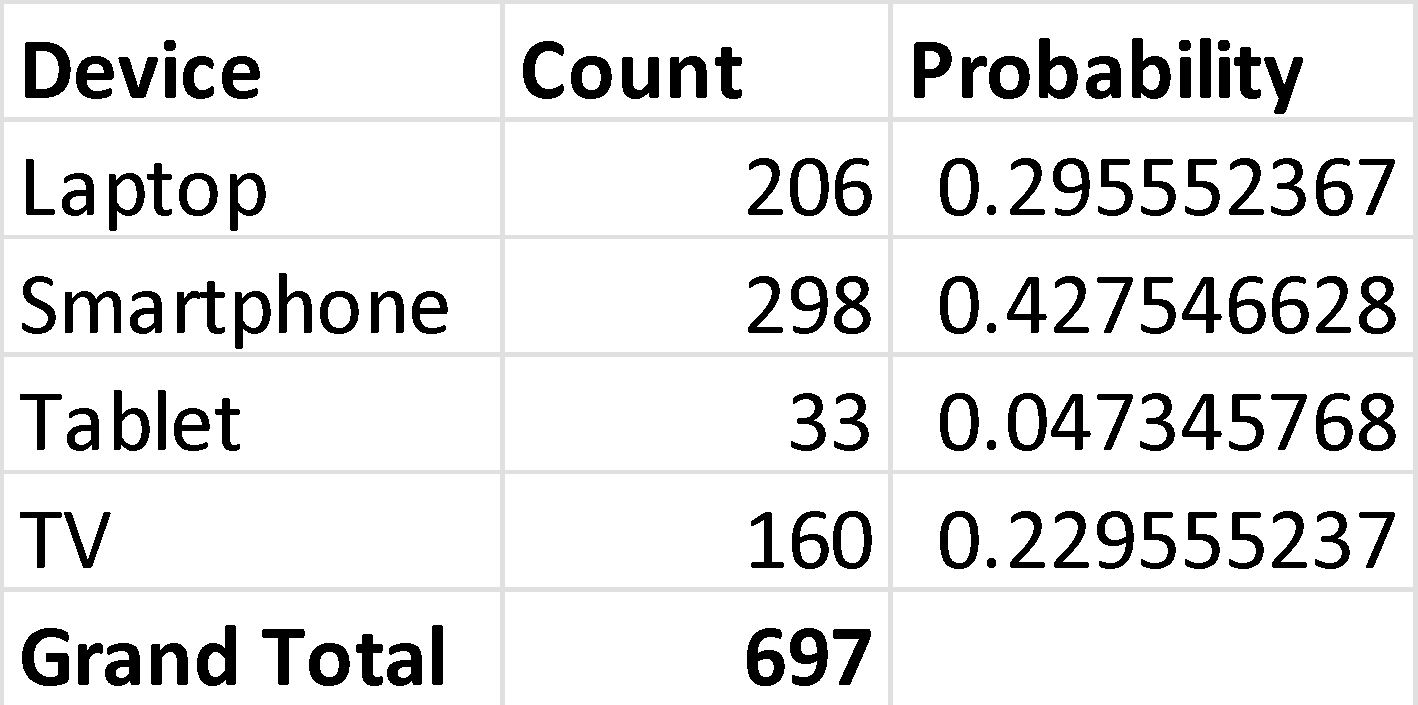


Following this distribution, we calculated the following probabilities:

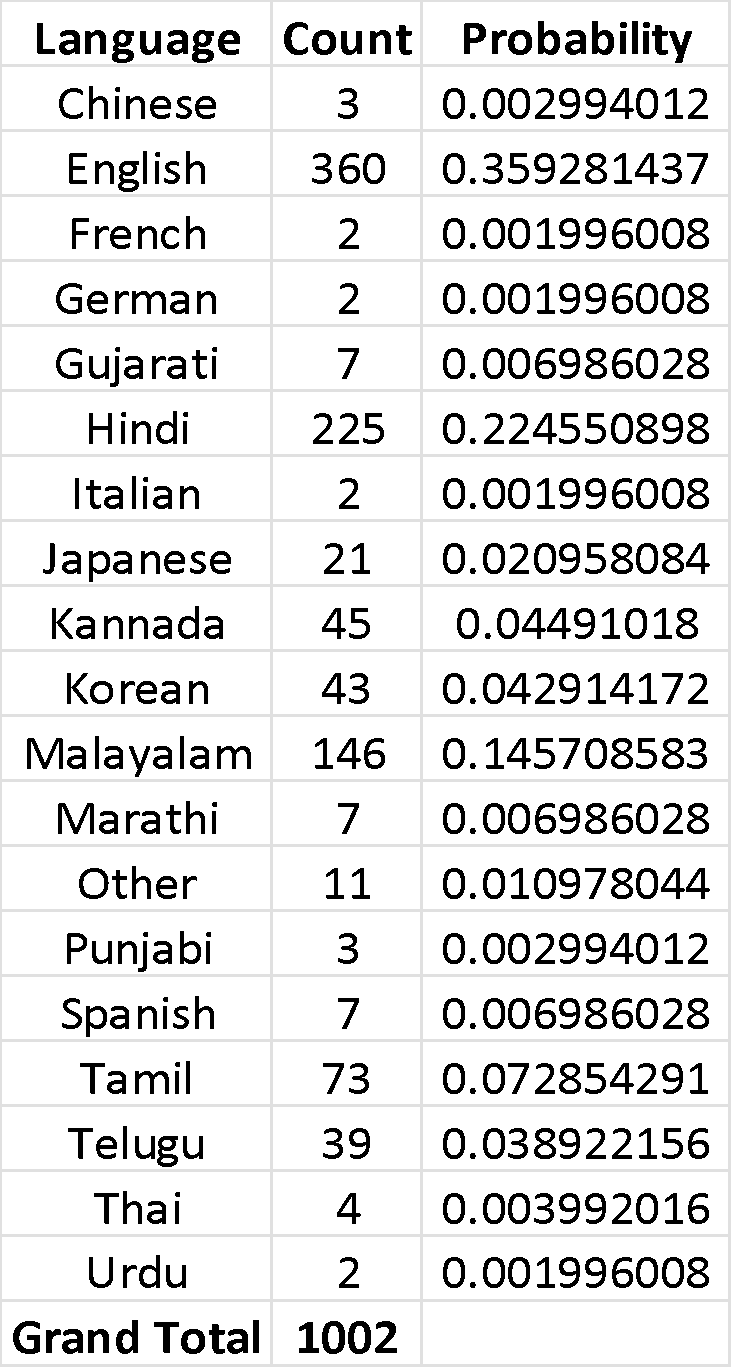
**1. In a random sample of 13 people, what is the probability that each person chooses a unique genre?**



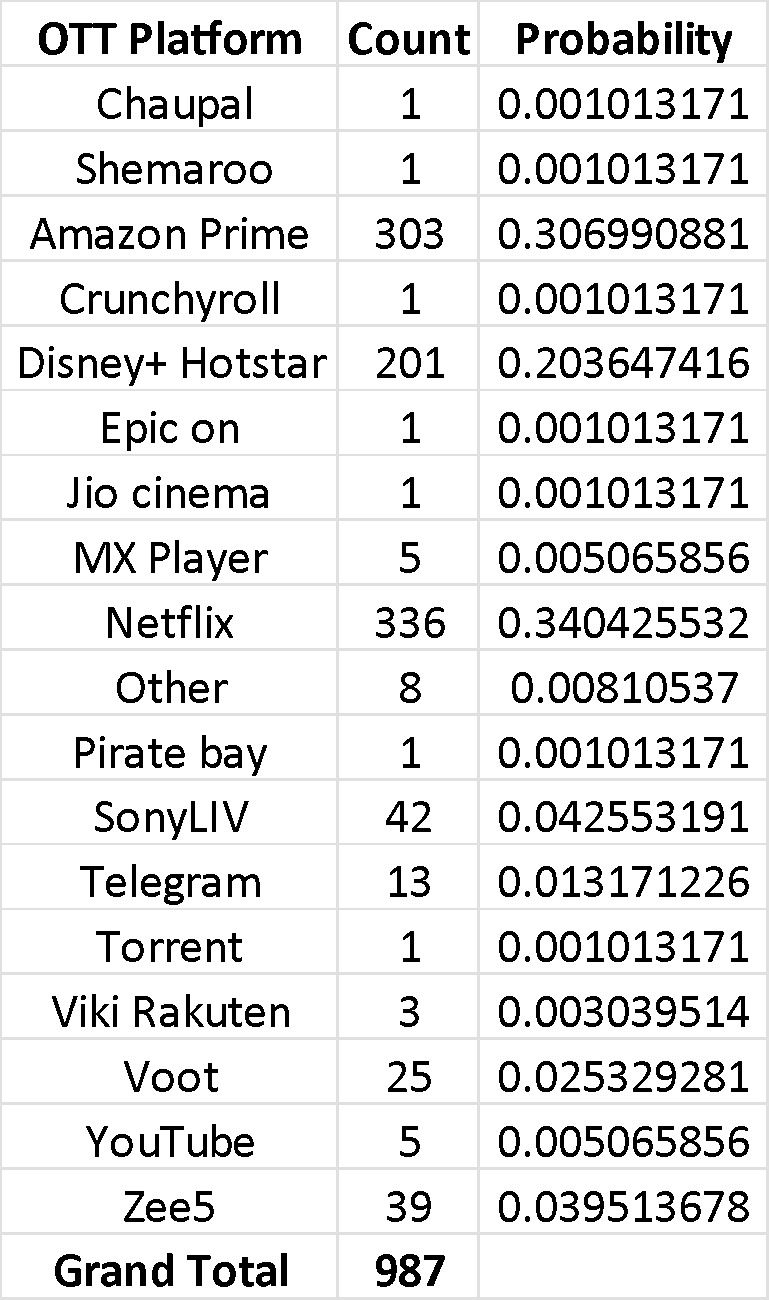
**2. Out of 10 random people selected, what is the probability that 4 choose smartphone, 3 choose laptop, 2 choose TV, and 1 chooses tablet?**



**3. What is the probability that 5 random people choose the 5 most popular languages each?**



**4. What is the probability that 3 people choose the top three most famous OTT platforms each?**



**User Interface Development:**

Our dashboard was created using the Python library called Plotly Dash. Plotly Dash is used to build interactive and responsive web-based dashboard applications.

Dash is specifically used to work with problems related to in-depth data analysis and visualization. For that, Plotly offers a wide range of 2D and 3D graphs which can be used in Dash.

Official Plotly documentation and official Dash documentation were referred in the process of creating the dashboard.

The IDE we used for coding is Pycharm.

**Summary and Recommendations:**

* The three south Indian states- Karnataka, Tamil Nadu, Kerala prefer watching content in their own mother tongue. So, developing the platform entirely to the regional language, or simply adding more local language content could be the way to go in these three states.
* The top three genres preferred by people are- Action, Drama and Comedy. So, OTT platform developers should include more movies/shows of these genres.
* The most common device used was smartphone. Hence, OTT platforms should offer more smartphone friendly (budget) subscriptions.
* According to our data, most people belong to town/city areas. So, increasing the service in those areas would benefit the customers as well as the company.
* Maximum population of OTT watchers are students. Hence, it would make more sense if more content targeted towards this age group was made.
* English was chosen to be the most preferred language by maximum people. This indicates a demand for international content in the Indian society.