## Analysis Assignment 1 - Part B

##### **Group 6:** Sara Hurley-Worthing, Ishaan Kapoor, Zaida Pacheco, Melissa Tverberg

#### ETL Process:

Based on the feedback provided from Part A of the assignment, we decided to focus on six relevant questions that can be analyzed using the provided data. In the ETL process, we linked each source of provided data by copying each data file into its own tab in an Excel spreadsheet to create a mini database; which in this case is also a flat file, attached below.

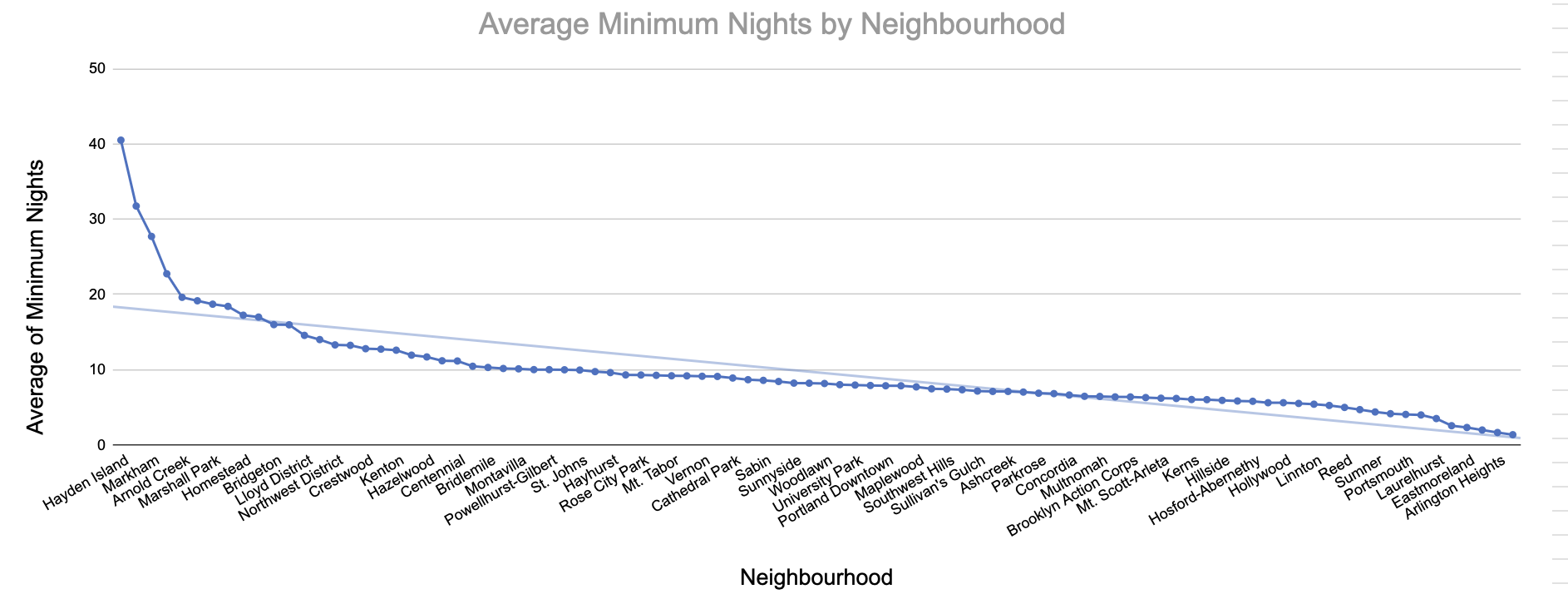
Our first step in cleaning the data was to review columns and rows to assess what data was there and how it could be altered. For the “AirBnB Dec 2023 – Listings” data for example, we added columns and populated them using the LEFT, RIGHT, MID, and VLOOKUP functions. By cleaning a column’s data down to a single word, we isolated the Type of Rental, Portland District, and Neighborhood available in the data. As we created visual representations of data, further cleaning was needed, for example using the CLEAN function to remove spaces. Another example is filtering the Rental Type column to pair down the number of filters to better analyze, cross checking the data against other columns to confirm accuracy, to assess if the rental was a shared space or entire property. It was also important to “normalize” the data by capitalizing column and row headings, choosing one name for duplicates, eg “2 bedroom” v. “2 bedrooms”. While cleaning the data we chose to include rows with missing data to compare volume of data against how much was missing, as we observed missing data is a material amount in several of our questions.

Next, we were able to sort and filter the data into Pivot Tables. During the process of transforming the data and utilizing Tim’s feedback on our original questions, we decided to simplify the questions we could analyze without needing any additional data. Finally, once we had isolated our core six questions, we created a Pivot Table for each question, comparing only two or three ideas at a time, such as the minimum count of days each Type of Rental requires, broken down by the number of bedrooms.

#### Analysis:

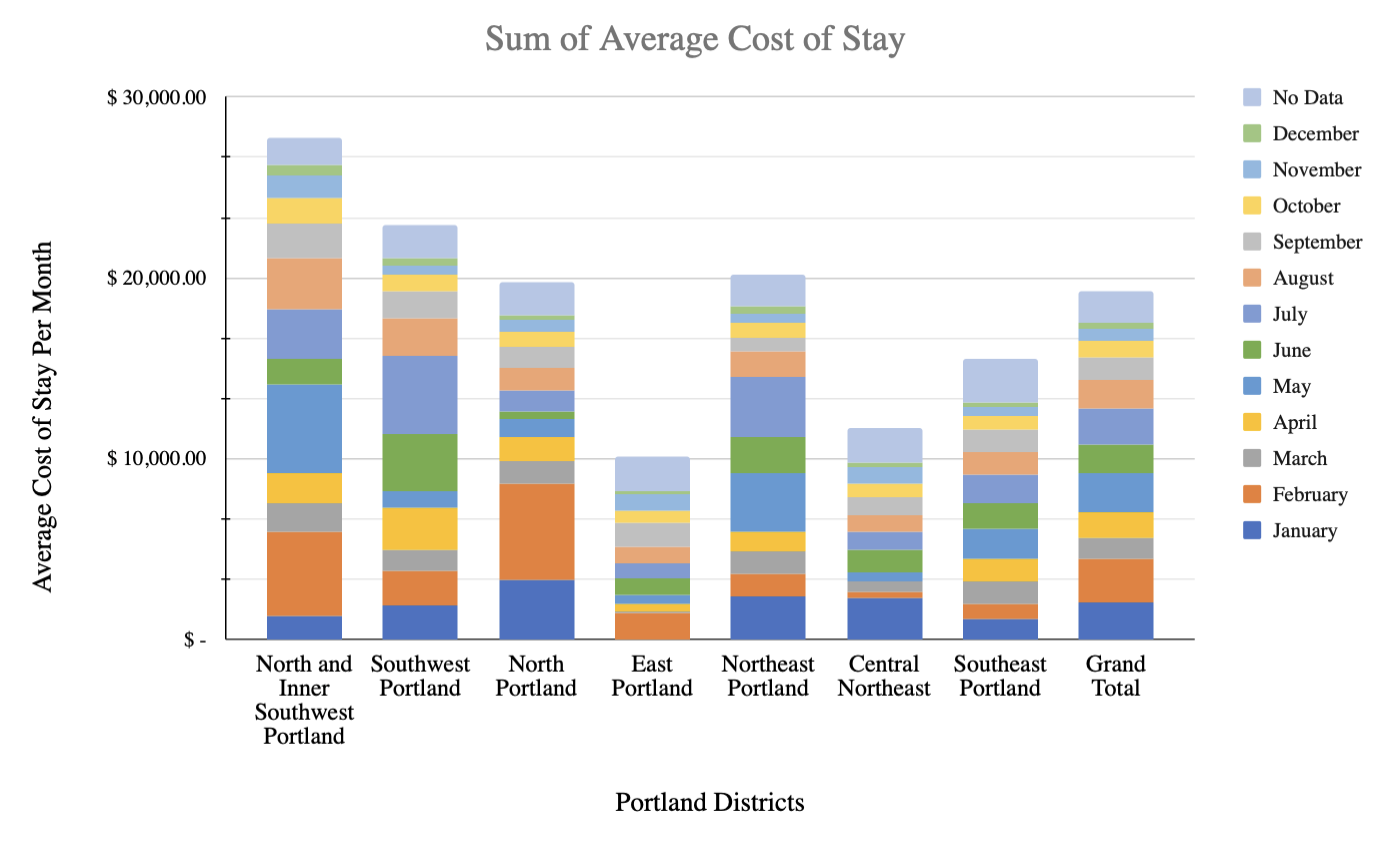
**Question 1:** **What is the average number of days stayed in each Portland neighborhood? And what is the count of stays per month for each neighborhood?**

The average number of days stayed in each neighborhood is about 10 days. The Hayden Island neighborhood has the highest minimum average at just over 40 days, and the lowest is Arlington Heights at 1.7% days. Neighborhoods predominately have 2 and 1 day minimum stays, making up 33% and 31% of the total city. Data is from the Clean AirBnB Dec 2023 – Listings and Portland Neighborhood. The three most-stayed at neighborhoods are Overlook, Pearl, and Northwest District, with occupancy at an average of 283 days of the year.



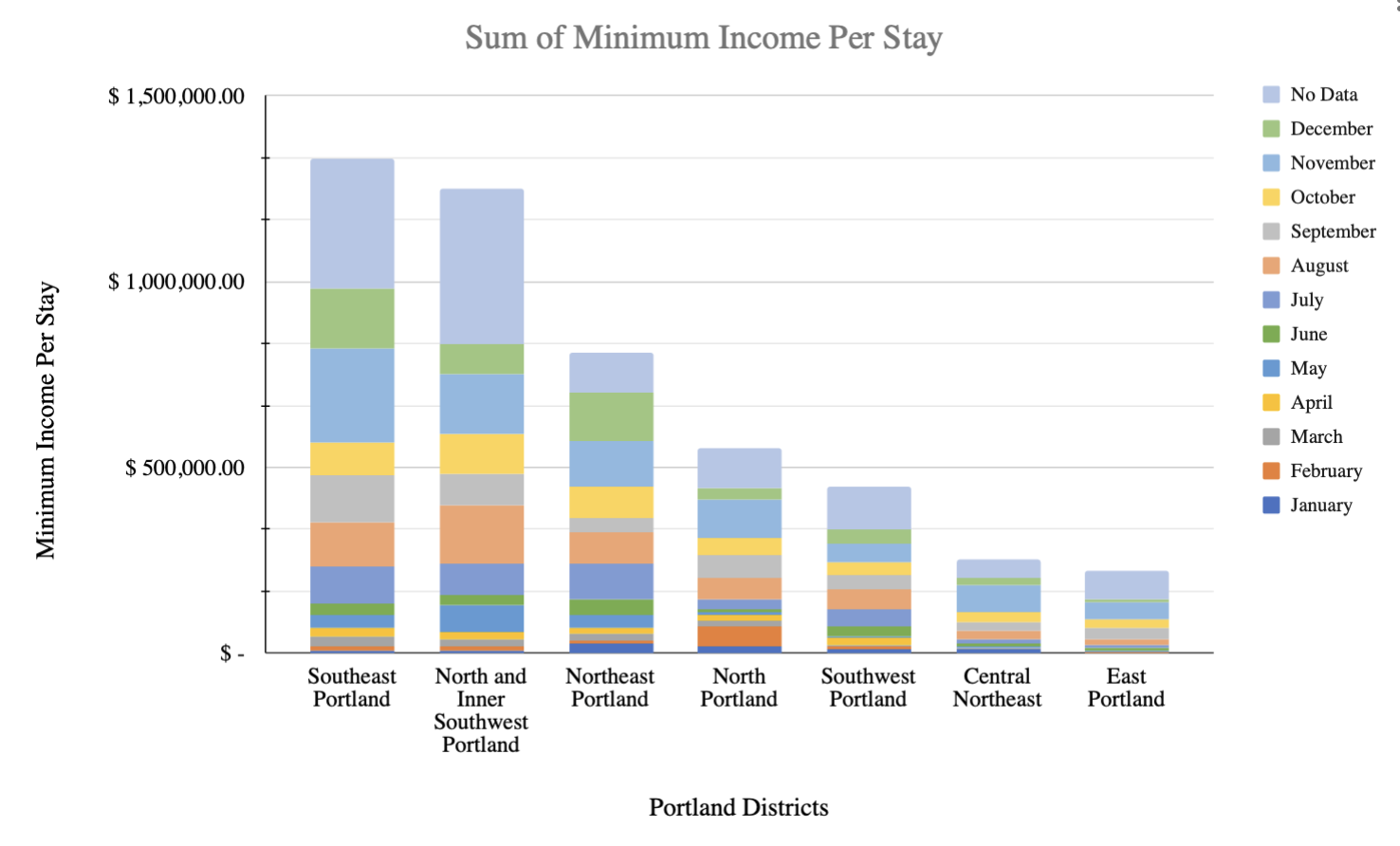
**Question 2**: **What is the average cost of rent per day in each Portland district by month?**

In analyzing Portland using neighborhoods v. districts, we found districts provided more holistic information, as there are only seven districts in the data, compared to 92 neighborhoods. Neighborhoods vary in size, whereas the districts are more similar in size. The average cost of rent per night stay is about $115. And the highest earning months are April, June, and October, making up 34% of the yearly income. The data used is from the Clean AirBnB Dec 2023 – Listings and Portland Neighborhood tabs. The overall highest earning districts are North and Inner Southwest, Southwest, and Northeast Portland. The North and Inner Southwest Portland district has the highest overall cost per stay at $250 per night on average, with an average minimum cost per stay of $1,200. The North and Inner Southwest district’s most profitable months are September, February, and June. The East Portland district has the lowest average cost per stay at $115, and the least expensive district overall, making about $785 a month on average. It is also noteworthy to state missing data makes up a significant portion of the overall data, and it is worthwhile to gain access to additional data.

****

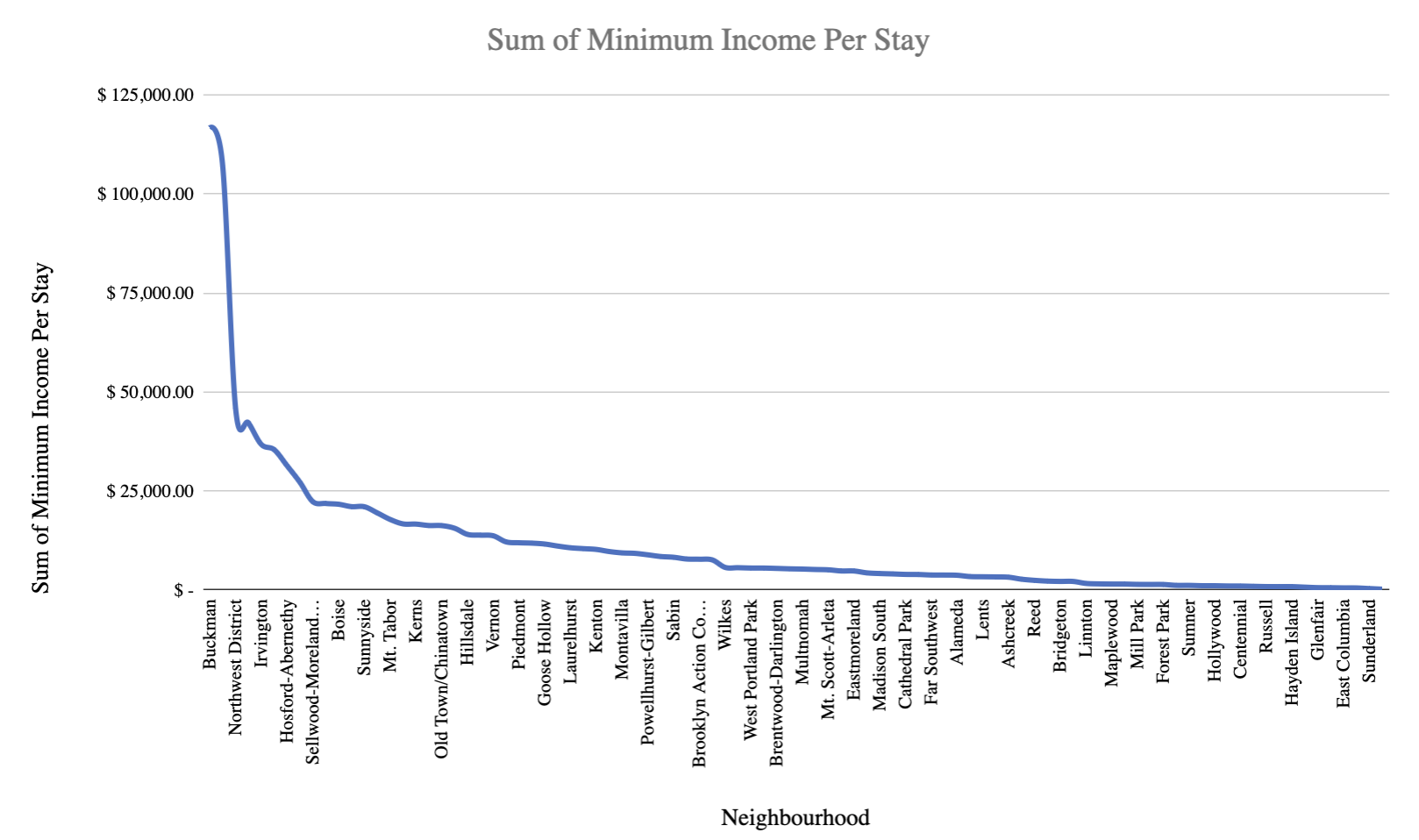
**Question 3: Which districts make the most income per night? Overall monthly income?**

The Southeast district is the highest earning in Portland, at $1.33M per year on average, with the North and Inner Southwest district at a close second with $1.25M per year on average. The highest earning months are October and November, making up a whopping 42% of the year. East Portland comes in last at about $200k per year on average. It is worth noting that missing data represents a material portion of the provided data and it would be useful to procure additional data, especially when the two highest earning districts have at least 20% of overall data missing.

****

**Question 4: Which neighborhoods are the most popular for (short-term) renters?**

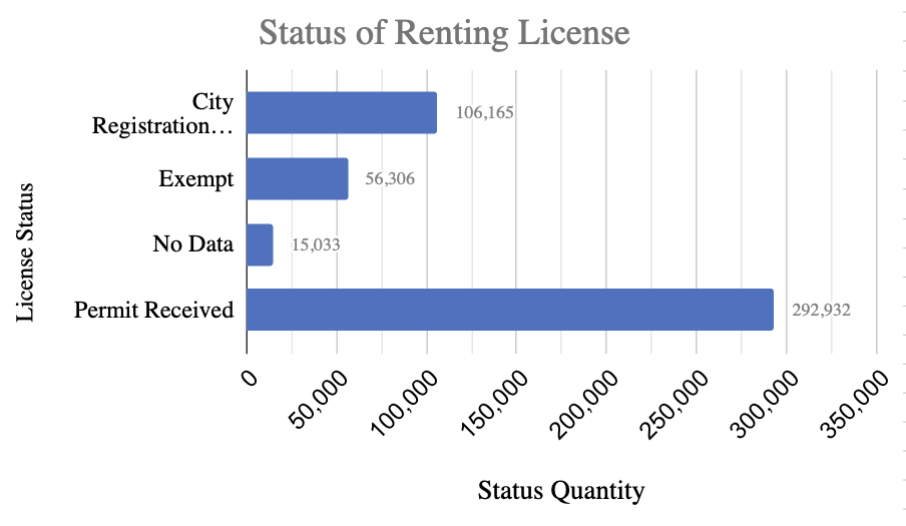
Our analysis focused on identifying popular short-term rental areas. After careful data collection and analysis, we created a comparative pivot table to assess these areas based on their rental costs. The resulting chart revealed that Buckman, Northwest, and Irvington were the most favored destinations for short stays, particularly among affluent individuals. Additionally, our analysis indicated minimal variation among other locations beyond the top five, emphasizing distinct preferences among short-term renters. Overall, our findings highlight Buckman's prominence among higher-income earners and provide valuable insights into rental preferences for short-term tenants.



**Question 5: Can I rent with a pending permit?**

After thoroughly cleaning and scrutinizing the data, we generated a pivot table comprising information on license status and the corresponding number of licenses. From this analysis, we deduced that out of the total 470,436 applicants, approximately 3.1% had missing data, indicating a small portion where information was not available.

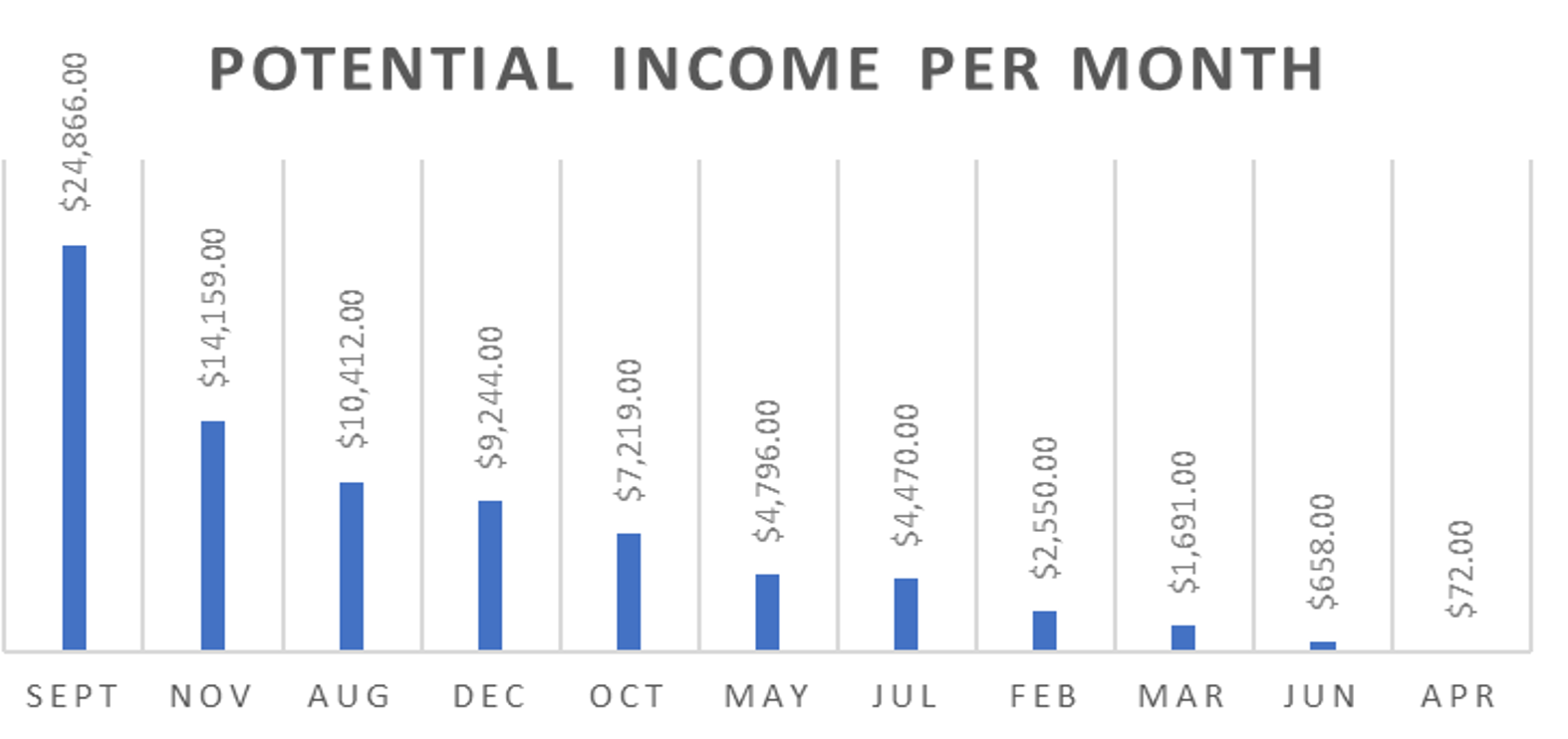
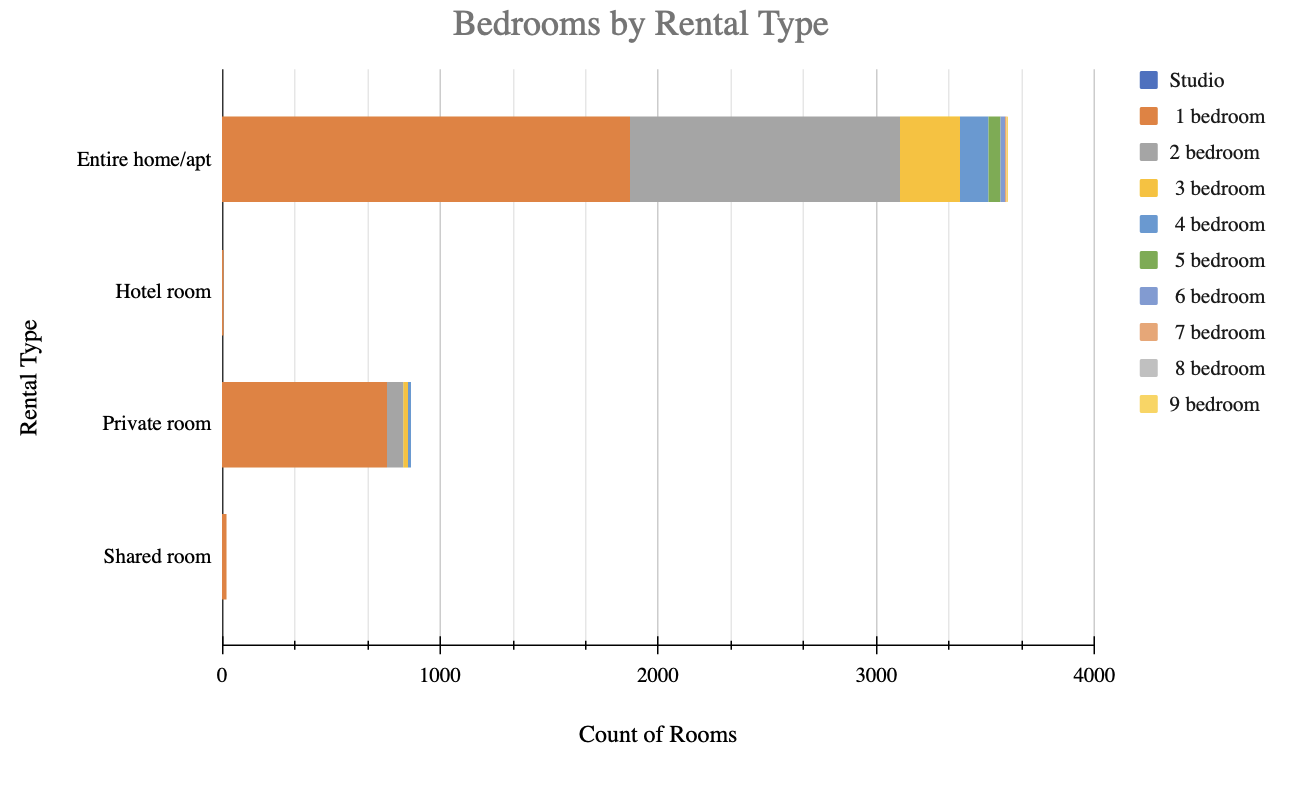
Breaking down the license status further, it was observed that approximately 62.26% of applicants obtained a renting license permit, indicating a significant majority who successfully acquired the necessary permits. Around 22.56% of applicants were found to be in a pending status for a renting license permit, suggesting that a considerable portion of applicants were awaiting approval. Additionally, roughly 11.96% of applicants were exempted from needing a renting license permit, indicating a notable minority who were not required to obtain this permit for their rental activities. These findings underscore the importance of regulatory compliance and the diverse landscape of rental licensing within the studied population of 470,436 applicants.

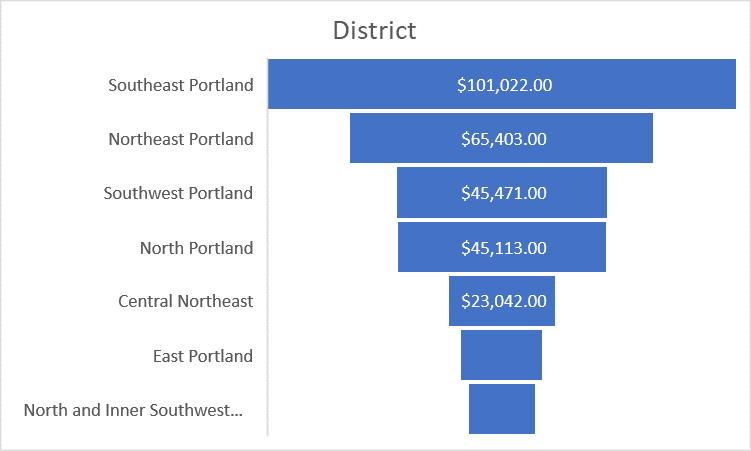
****

**Question 6: What size would you like the property to be? Number of bedrooms/bathroom, sq footage, etc. What would be the ideal rental? What are people looking for 1-2 bedrooms, home, studio, # bathrooms.**

The analysis for question 6 was focused more on type of property and number of bedrooms. This analysis excludes sq footage as we did not have this information available.

It appears that based on the analysis we did, people are more inclined to rent an entire home with 1,2 and 3 bedrooms in the Southeast Portland area. The second most appealing per analysis would be a private house with 1 to 2 bedrooms. The next type of rental is a private room/studio. The potential income on the entire home based on the minimum rental requirement was estimated at around $100K per year. How we came up with this information. We gather all the data from the “Clean AirBnb Listing” and clean-up by extracting from the “name” field and identifying “home type”. That is how we created categories for the rental type. We also added a field called “district” by using the V-lookup is how we were able to group the neighborhood groups into districts. The average income was calculated by adding a few fields and calculating the minimum night stay by the price night. After cleaning and creating a single source of truth we then created various pivot tables.



**

*Data Sources/Resources:*

* ***Tim’s*** [***Feedback From Part A***](https://canvas.pdx.edu/courses/82114/discussion_topics/867685)
* [AirBnB Dec 2023 - Listings](https://canvas.pdx.edu/courses/82114/files/9926715?wrap=1)
  + [Download AirBnB Dec 2023 - Listings](https://canvas.pdx.edu/courses/82114/files/9926715/download?download_frd=1)
* [AirBnB Dec 2023 - Reviews](https://canvas.pdx.edu/courses/82114/files/9926716?wrap=1)
  + [Download AirBnB Dec 2023 - Reviews](https://canvas.pdx.edu/courses/82114/files/9926716/download?download_frd=1)
* [Portland Neighborhood Associations](https://canvas.pdx.edu/courses/82114/files/9926707?wrap=1)
  + [Download Portland Neighborhood Associations](https://canvas.pdx.edu/courses/82114/files/9926707/download?download_frd=1)
  + ([portlandoregon.gov Links to an external site.](https://www.portlandoregon.gov/civic/index.cfm?c=28385))
* [Portland Neighborhoods - 2020 Data Profiles](https://canvas.pdx.edu/courses/82114/files/9926708?wrap=1)