# PROJECT
# PORTFOLIO

Prepared By:

# PROFESSIONAL BACKGROUND

I am a multi-talented professional with a diverse background in economics, education, and email copywriting. I earned a Bachelor of Science degree in Economics and honed their skills as a teacher for five years, fostering critical thinking and problem-solving in students.

With a passion for writing, I transitioned into email copywriting and has since become an expert in crafting compelling, persuasive emails that drive conversions and sales. In addition, I possess excellent skills in data entry, Microsoft Excel, Google Sheets, and Microsoft Word, enabling me to analyze and organize data effectively.

The unique blend of analytical and creative skills, combined with strong work ethic and attention to detail, makes me a valuable asset in any data analysis project. And I thrive in collaborative environments and am committed to delivering high-quality work on time and within budget.

Whether it's interpreting data trends or crafting engaging email content, I am dedicated to achieving outstanding results and exceeding expectations.

# TABLE OF CONTENTS

# UDEMY PROJECT DESCRIPTION

The aim of this project is to gain a better understanding of the pricing and subscription data for Udemy courses in four different subjects. By analyzing the data, we hope to identify trends and patterns that will provide insights into the behavior of Udemy's users.

Through this analysis, we plan to determine which subject areas have the most demand for courses and which areas may need more courses to be created to meet the needs of users. We will also create targeted strategies to increase revenue for the company by identifying pricing and subscription trends.

Additionally, we will identify which courses should have more free beginner courses to attract more users and increase enrollment. We will analyze if the duration of the course impacts the pricing strategy, and determine if shorter or longer courses are more attractive to potential customers.

The results of this project will provide valuable insights into the behavior of Udemy's users and inform the company's strategy for creating and pricing courses. This will ultimately lead to a more successful and profitable business model.

# THE PROBLEM

## <u>Defining the Problem:</u>

To better understand the Udemy course data and identify trends and patterns, we need to answer questions that seek to understand various aspects of the pricing and subscription data of Udemy courses for four different subjects.

Specifically, the aim to uncover information about the total number of subscribers in each subject, the variations in average content duration, price, and number of students across different subjects, the number of free and paid courses for each subject, the average price of web development courses at different levels, the 20 most popular courses and their attributes such as level, free or paid status, beginner course status, and duration, and the impact of content duration on the course price.

By answering these questions, the data analysis project can identify trends and patterns that can inform decisions about creating more courses, creating targeted revenue strategies, and identifying which courses should have more free beginner courses.

# DATA DESIGN

The aim of this study is to identify revenue increase, opportunities, and performance of Udemy courses across four subjects: Web Development, Graphic Design, Business Finance, and Musical Instrument.

The following data analysis design will be implemented to achieve the study's objectives:

1. Data clean up: The data will be cleaned up by removing duplicates, removing blank cells, and finding and replacing words where necessary. The headers will also be cleared for easy analysis.

2. Pivot Tables using Google Sheets: Pivot tables will be used to summarize and analyze the data. This will help to identify the revenue generated by each subject and provide insights into the performance of each subject.

3. Tableau for Visualization: Tableau will be used to create visualizations that are suitable for large datasets and detailed analysis. The visualizations will help to identify trends and patterns in the data and provide insights into revenue increase and opportunities.

4. Analysis: The data will be analyzed to identify the revenue generated by each subject, the average content duration/price/number of students for each subject, and the number of free and paid courses for each subject. The analysis will also show if content duration impacts the price of the course and identify the 20 most popular courses, including their level, whether they are free or paid, whether any are free beginner courses, and the duration of the courses.

Based on the analysis, it is expected that Web Development will have the highest revenue and better performance than other subjects. The insights gained from this analysis will provide valuable information for targeted strategies to increase revenue and identify opportunities for growth in the Udemy platform.
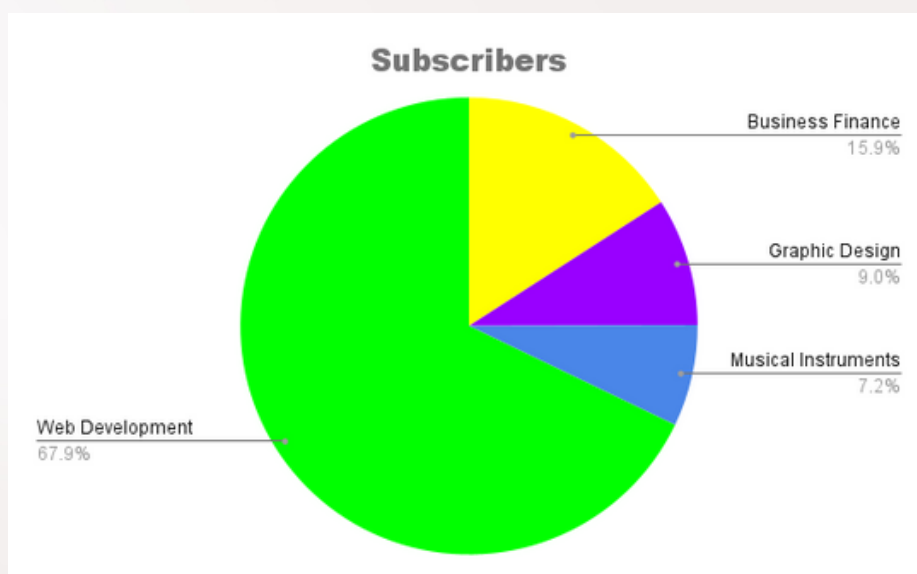
# FINDINGS

## Finding 1:

In the Udemy dataset of four subjects, the total numbers of subscribers in percentage for each subject are 7.2%, 9.0%, 15.9%, 67.9% respectively as seen in the chart.
Web Development has the highest number of enrolled subscribers with 68% with a total number of 7,981,935 out of 11,760,483 subscribers.

## Table 1 :
**Total number of subscribers in each subject.**

| subject | SUM of num_subscribers |
|---|---|
| Business Finance | 1868711 |
| Graphic Design | 1063148 |
| Musical Instruments | 846689 |
| Web Development | 7981935 |
| **Grand Total** | **11760483** |

## Fig 1:



Subscribers

- Business Finance 15.9%
- Graphic Design 9.0%
- Musical Instruments 7.2%
- Web Development 67.9%

# Finding 2:

The analysis shows that Web Development has the highest average content duration, price and number of students across all subjects and business finance following with content and price but Graphic design takes over it in subjects while musical instruments has the lowest of all four subjects.

## Table 2:

### Subjects vs Content Duration/Price/Subscribers.

| subject | AVERAGE of content_duration | AVERAGE of price | AVERAGE of num_subscribers |
|---|---|---|---|
| Business Financ | 3.62 | 69.08 | 1617.93 |
| Graphic Design | 3.65 | 58.77 | 1820.46 |
| Musical Instrume | 2.89 | 49.65 | 1265.60 |
| Web Developme | 5.59 | 77.04 | 6635.02 |
| **Grand Total** | **4.14** | **66.46** | **3256.85** |

## Fig 2:

# Findings 3:

In the table below, it shows that all four subjects are represented by free or paid which shows web development and business finance with the most free(134 & 96) or paid (1,069 & 1,059) courses while graphic design and musical instruments have the lowest free (35 & 46) or paid (594 & 623).
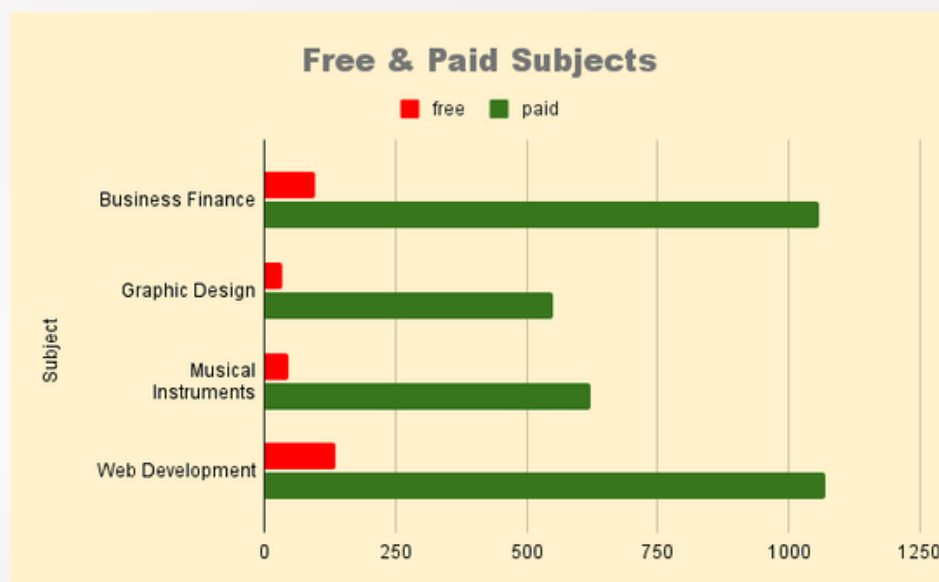
This makes web development the highest free or paid subject for all courses as represented in the chart below.

## Table 3:

**Subject vs Free or Paid**

| subject | free | paid | Grand Total |
|---------|------|------|-------------|
| Business Finance | 96 | 1059 | 1155 |
| Graphic Design | 35 | 549 | 584 |
| Musical Instrume | 46 | 623 | 669 |
| Web Developme | 134 | 1069 | 1203 |
| **Grand Total** | **311** | **3300** | **3611** |

## Fig 3:



Free & Paid Subjects

# Findings 4:

The area chart below shows the total average price of web development at different levels; All levels has 74.55, Beginner Levels has 78.54, Expert Level has 67.14 and Intermediate Level has an average of 85.07 which is the highest.
Which makes Expert Level the course with the lowest price of all the courses in Web Development.
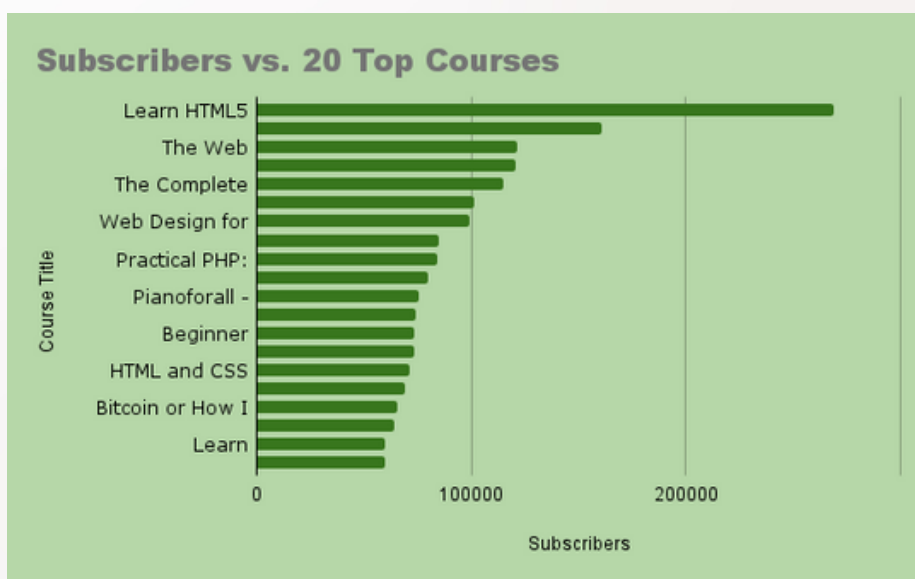
# Fig 4:
**Level vs Average Price**
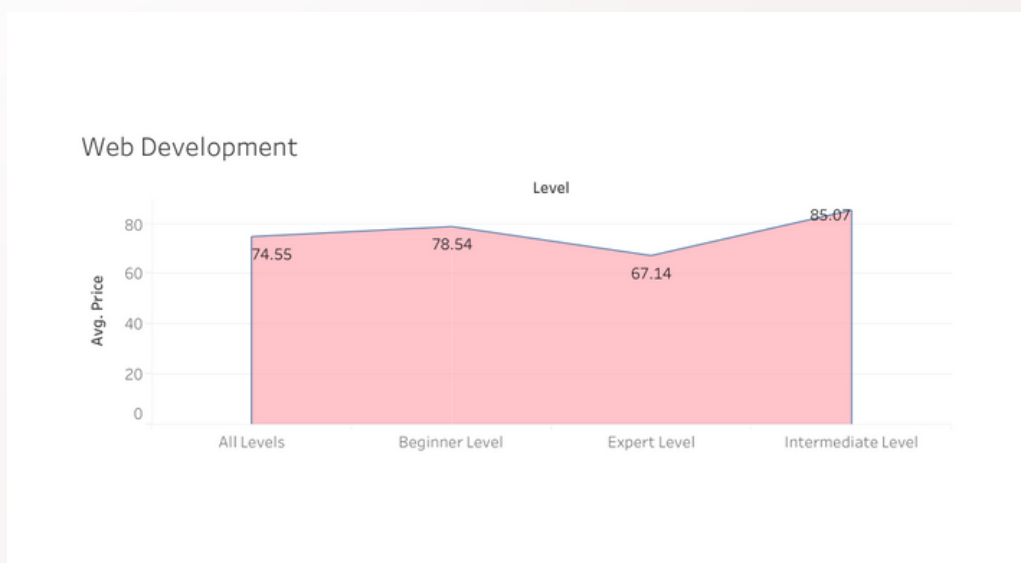
# Findings 5:

 In the column chart above, it shows the 20 most popular courses of all the four subjects while the first 5 is from Web Development, which makes it the most enrolled subject with the highest number of subscribers on all levels whether free or paid, free beginner courses and course duration.

## Fig 5:
 **Top 20 Courses**

# Findings 6:

In the scatter plot below it shows Content duration (0-80) vs Price(0-200). Looking at the price on the graph from 0 to 200, you will see that at zero price it is free and not the one with the lowest content duration. Also the content duration which is between 40 & 60 has the highest content duration close to 80 followed by the price at 200.
Which concludes that content duration does not impact the price of the courses.

# Fig 6:
**Content Duration vs Price**



Content Duration vs Price
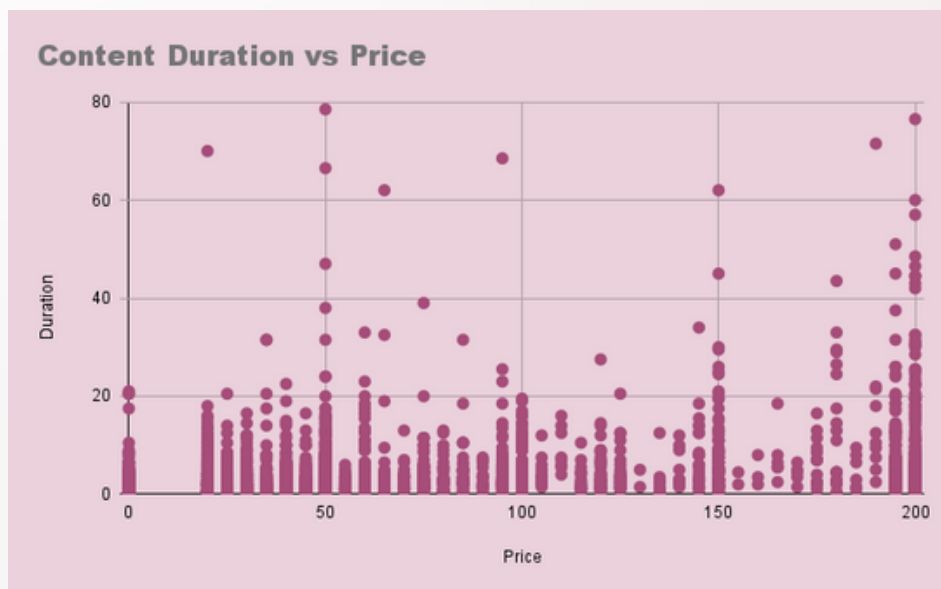
# SUMMARY

Based on the data analysis, it is evident that Web Development is the most popular and revenue-generating subject on Udemy. It accounts for almost 68% of the total subscribers, and the top 5 courses are also for Web Development.

In comparison, Business Finance has almost the same number of total courses as Web Development, but Web Development has almost double the content duration and four times the average number of subscribers. This suggests that Web Development courses have better performance and provide more revenue-generating opportunities than Business Finance.

Furthermore, the data shows that for all four subjects, there are more paid courses than free courses on all levels, and there are only a few beginner courses available. Among the four subjects, Web Development courses have the highest number of course prices, while musical instrument courses have the lowest.

Interestingly, the content duration does not affect the price of all four subjects. This means that longer courses do not necessarily mean higher prices, and Web Development courses with the highest content duration still have the highest price and number of subscribers.

Overall, the use of pivot tables in Google Sheets and the visualization in Tableau has helped identify trends and patterns in Udemy's pricing and subscription data for different subjects. The findings can be used to create targeted strategies to increase revenue and to identify opportunities for creating new courses.

# CONCLUSION

## Actions and Recommendation:

Based on the analysis of the four Udemy courses, the following actions and recommendations can be made:

- Increase investment in Web Development courses: As Web Development has the highest revenue and the highest number of subscribers, it is recommended to increase investment in this area by creating more courses and targeted strategies to further increase revenue.

- Consider creating more beginner courses: The analysis showed that there are fewer beginner courses in all four subjects. Creating more beginner courses can help to attract more subscribers, especially in subjects where there is a high demand but low supply of beginner courses.

- Create a community: This is where subscribers can meet to ask questions on their various subjects whenever they have a problem especially for courses with low subscribers. Which can also attract subscribers to take more subjects and increase revenue.

- Evaluate the pricing strategy: While there are more paid courses than free courses for all four subjects, the analysis showed that Web Development has the highest number of course prices. It may be worth evaluating the pricing strategy for Web Development courses to ensure that they are competitively priced and not discouraging potential subscribers.

# CAPSTONE PROJECT DESCRIPTION

Data Source Analysis Expected Result The data for this analysis will be sourced from a third-party website Kaggle, that provides scraped data on New York City Airbnb listings. This data includes information on listing characteristics e.g., location, room type, price, re and trends etc.

The analysis will utilize a combination of descriptive statistics to examine the prices and the neighbourhood, as well as the variation in prices by room type and neighborhood group and also identify trends in Airbnb listings over time.

We expect to find that prices and listing activity vary considerably by room type and neighborhood group with some areas having more listings and higher prices than others. We also anticipate observing a general upward trend in Airbnb listings across New York City, with some variations across different neighborhood groups and that certain room types will be more popular and have higher availability than others.

# THE BUSINESS PROBLEM

The problem is that Airbnb hosts in New York City are facing challenges with pricing, marketing, and customer experience due to the variation in prices by room type and neighborhood group, as well as pricing and popularity trends.

Additionally, the high demand for Entire Home/Apartment listings makes them less available, despite their higher cost, while Shared room listings have lower prices but are always available for listing. As a result, hosts may struggle to attract customers and optimize their profits in a competitive market.

The questions we seek to answer by analyzing the data are:

- What is the average price of an airbnb listing in each neighborhood and how does it vary by room type?
- Which neigbourhood has the highest and lowest average price?
- What is the percentage of listings per room type?
- Are there trends in the popularity of airbnb hostings in different neighbourhood over time?

# FINDINGS

## Finding 7:

In the table below, the average price for each neighborhood group is: Bronx (87.5), Brooklyn (124.4), Manhattan (196.6), Queens (99.5) and Staten (114.8). Its variations in room types, Entire Home/ Apartment is the highest in Manhattan at 249.2 while Bronx has 127.5 lowest, followed by Private room in Manhattan at 116.8 while Staten Island has 62.3, and Shared room in Brooklyn with the lowest at 50.5 while Manhattan has the highest at 89.0.

However, Shared room has the highest availability, while Entire Home/ Apartment is less available despite the high demand

## Table 4:

**Average price of Airbnb listings versus Room type in neigbourhood group.**

| | A | B | C | D | E |
|---|---|---|---|---|---|
| | AVERAGE of price room_type | | | | |
| | neighbourhood_ | Entire home/apt | Private room | Shared room | Grand Total |
| | Bronx | 127.5 | 66.8 | 59.8 | 87.5 |
| | Brooklyn | 178.3 | 76.5 | 50.5 | 124.4 |
| | Manhattan | 249.2 | 116.8 | 89.0 | 196.9 |
| | Queens | 147.1 | 71.8 | 69.0 | 99.5 |
| | Staten Island | 173.8 | 62.3 | 57.4 | 114.8 |
| | **Grand Total** | **211.8** | **89.8** | **70.1** | **152.7** |

## Fig 7:



Avg Price vs Neighbourhood(Room Type)
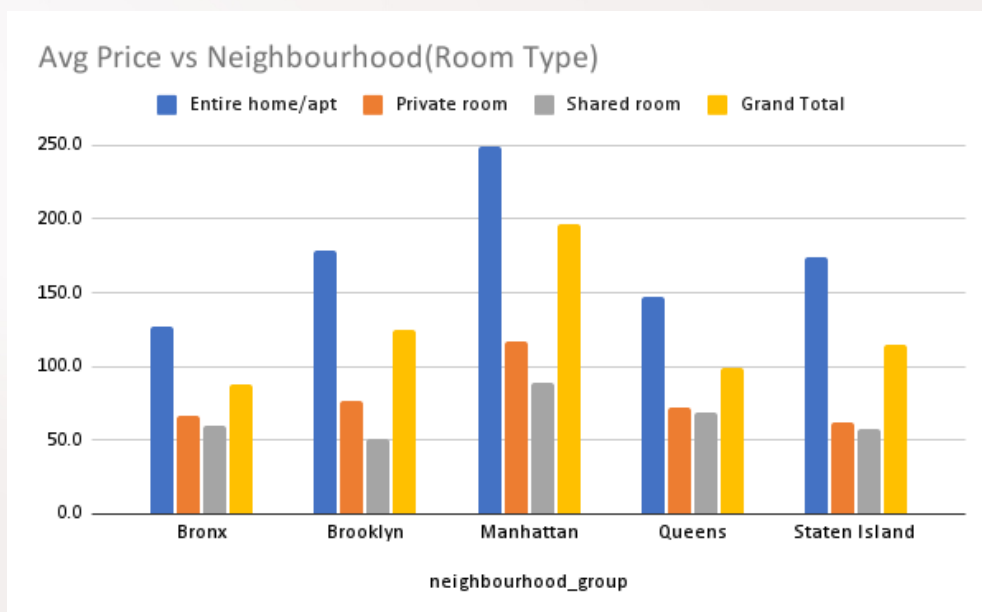
# Findings 8:

The results indicate that prices and listing activity vary considerably by room type and neighborhood group. Manhattan has the highest number of listings for and the highest average price at 196.9, while Bronx has the lowest number of listings and the lowest average price at 87.5.

## Fig 8:

**Neighbourhood vs Average Price**
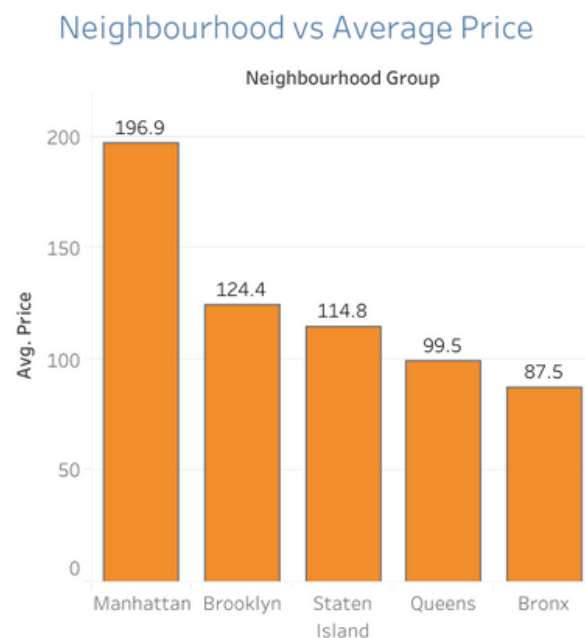
# Finding 9:

The percentage shown in the pie chart shows that, Entire Home /apartment has 52%, Private room has 45.7% and shared has only 2.4% of the listings per room type. The high demand for Entire Home/Apartment is less available even though it costs more unlike Shared home which has a low average price but is always available for listing.

## Fig 9:
**Percentage Of The Listing Per Room Type**
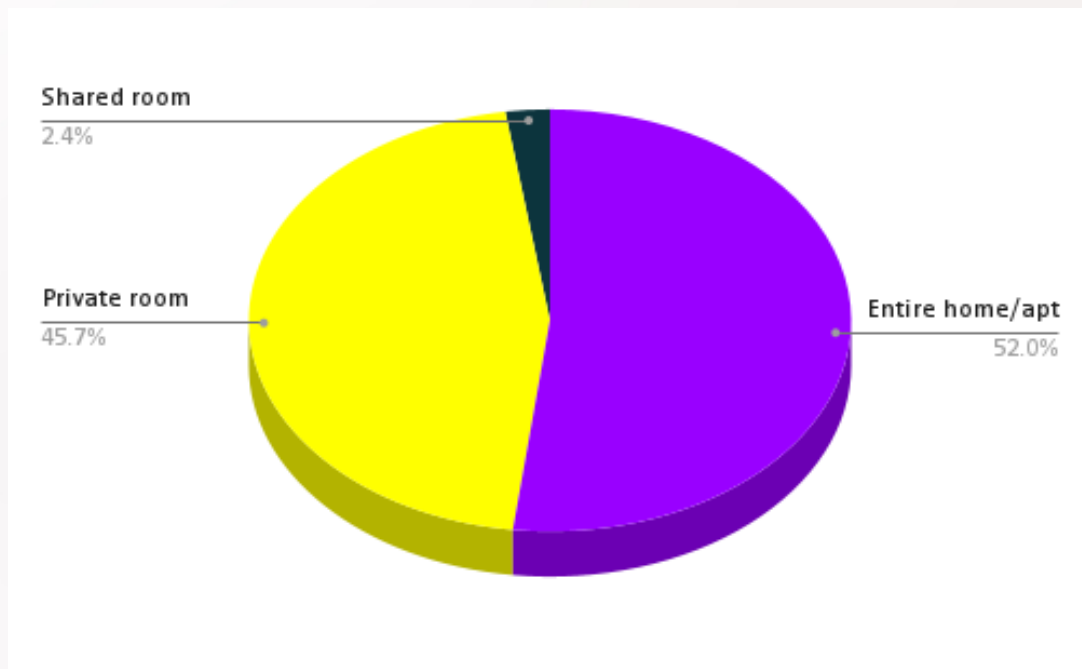
# Findings 10

The trend of Airbnb listings in New York City showed a slow increase from 2014 to 2016, with a slight increase from 2016 to 2017 and again in 2018. It peaked in 2019, Staten Island seem to have started airbnb listing between 2018 and 2019 and it had the lowest number of listings, with low availability despite the lowest average price.

## Fig 10:

### Trends of AirBnB Listing
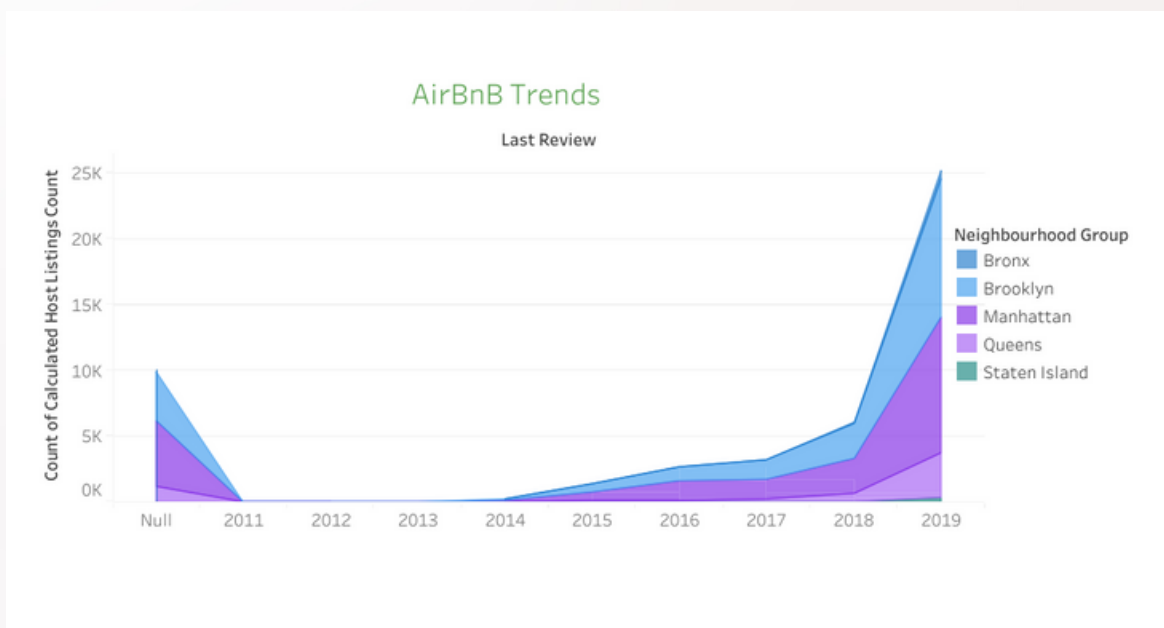
# SUMMARY

The data used for this analysis was obtained from Kaggle, providing information on Airbnb listings in New York City, including location, room type, price, reviews, host information, and availability data. Descriptive statistics were utilized to examine the variation in prices by room type and neighborhood group, the neighbourhood with the highest Airbnb listing price, percentage of price per room type and trends in Airbnb listings over time.

The results indicate that prices and listing activity vary considerably by room type and neighborhood group. Manhattan has the highest number of listings and the highest average price, while Bronx has the lowest number of listings and the lowest average price.

Also, the average price in percentage of Entire Home/ Apartment is the highest, followed by Private room, and Shared room with the lowest. However, Shared room has the highest availability, while Entire Home/ Apartment is less available despite the high demand.

In conclusion, the trend of Airbnb listings in New York City showed a slow increase from 2014 to 2016, with a slight increase from 2016 to 2017 and again in 2018. It peaked in 2019, and Staten Island had the lowest number of listings, with low availability despite the lowest average price.

# ACTIONS AND RECOMMENDATION

Based on the analysis of the Airbnb data in New York City, we can make the following recommendations:

- Price Optimization: The analysis shows that prices vary significantly by room type and neighborhood group. Hosts can use this information to optimize their pricing strategy by setting competitive prices for their listing in their respective neighborhoods and room types. Hosts can also adjust their prices based on the availability of their listings to maximize their revenue.

- Marketing and Customer Experience: The data indicates that Manhattan and Brooklyn are the top neighborhoods for Airbnb listings, and that certain room types are more popular and have higher availability than others. Hosts can leverage this information to improve their marketing efforts by targeting potential guests in popular neighborhoods and highlighting the features of their listings that make them stand out. Hosts can also focus on enhancing the customer experience by providing top notch service and amenities to earn positive reviews and increase the popularity of their listings.

- Room Type Optimization: The analysis shows that Entire Home / Apartment is the most popular room type with 52% of the listings, while shared rooms have only 2.4% of the listings. Hosts can use this information to optimize their room types by focusing on offering more entire home/apartment listings to meet the demand. Capstone Project Hosts can also consider offering shared room listings at competitive prices to increase their availability and appeal to budget-conscious travelers.

- Trend Analysis: The trend analysis indicates that the demand for Airbnb listings in New York City has been steadily increasing since 2014, with a peak in 2019. Hosts can leverage this information to make informed decisions on their Airbnb business by monitoring trends in their respective neighborhoods and adjusting their pricing and marketing strategies accordingly.

# APPENDIX

## Google SpreadSheet Data Set for Udemy Project

https://docs.google.com/spreadsheets/d/1a-qZyBD5YfmEZvmrMYR8Xvd7b84sClk930RhnTrmf-w/edit?usp=sharing

## Tableau Visualization For Udemy Project

## Google SpreadSheet Dataset for AirBnb Analysis (Pivot tables and visualization)

https://docs.google.com/spreadsheets/d/15FUHhRnK1r9XppNX5Y-wBKl75hAuCaWRW7oEN11G3mE/edit?usp=sharing

## Kaggle for NYC Airbnb Data

https://www.kaggle.com/datasets/dgomonov/new-york-cityairbnb-open-data?resource=download

## Tableau for Visualisation of NYC AirBnB