DATA ANALYSIS PORTFOLIO

PREPARED BY

EntryLevel_ Data_Portfolio
_Noname



Professional Background

I am a recent graduate from the University of X. I majored in Applied Mathematics and Computer Science. I have knowledge in the following programming languages C++, Java and Assembly Language.

I have been exposed to working with microsoft excel, google sheets, sql, tableau and planning to acquire skills working with Microsoft PowerBI, R and Python.

Portfolio Outline

Professional Background	1
Table of Contents	2
Udemy Project Description	3
The Problem	4
Design	5
Findings	6
Analysis	8
Conclusion	9
Air Bnb NYC 2019 Project Description	10
Data Design	11
Findings	12
Data Analysis	13
Conclusion	14
Appendix (optional)	15



Udemy Project Description

Data from four Udemy subjects that are being offered online was given, a course revenue had to be calculated from the different levels at which all the data was being given to its number of subscribers.

The purpose of this report was to allow consolidation of data to take place, followed by the cleaning of the data. The objective was to find ways to share light with Udemy of where they could change how they do things inorder to change increase their current revenue with the current available statistics.

The methods used was consolidating the different datasets into one sheet using Google sheets. I then had to clean the consolidated data to avoid working with repeating data. Pivot tables and Vlookup was created to see different patterns that the data followed.

From the data I then analysed that Udemy is offering more skills and time on courses that are being given for free and charging a lot on those few ones that are being given at a price. So it would be best if they forcused more on offering courses that are being paid and highly rated for more revenue.

The Problem

- Udemy is looking at ways to increase their revenue so they dont have a clear understanding of their subject performace.

- The project had a time frame of about 3weeks.

- Data collected should be from the four subjects of interest which are Business, Graphic arts, Music and Web Development. It should be [presented as one set of data that allows one to make clear comparison among the subjects.

- The question that can be asked is how much changes are they willing to do inorder to reach their goal of increasing revenue.

Data Design

Four subject data for Udemy has been consolidated and merged into a single batch of data in which repeating data has been cleared. Google sheet was used to clean the data and add more sections that were needed for future comparisons.

Vlookup has been used in order to calculate the top 20 performing subjects as well as pivot tables to compare data virtually. Tableau has been used to represent the virtualisation of the data as it made things easier for looking at trends made by the data .

Findings

figure 1: average number of subscr vs paid/free/subject

AVERAGE of num_subscribers	free/paid		
subject	Free Course	Paid Course	Grand Total
Business Finance	6283.625	1155.692237	1569.026868
Graphic Design	8137.742857	1372.710758	1766.026578
Musical Instruments	6624.673913	854.8170347	1245.130882
Web Development	17863.93284	5227.472404	6635.024938
Grand Total	11532.31833	2429.10312	3199.260881

figure 2: average number of subscr per subject

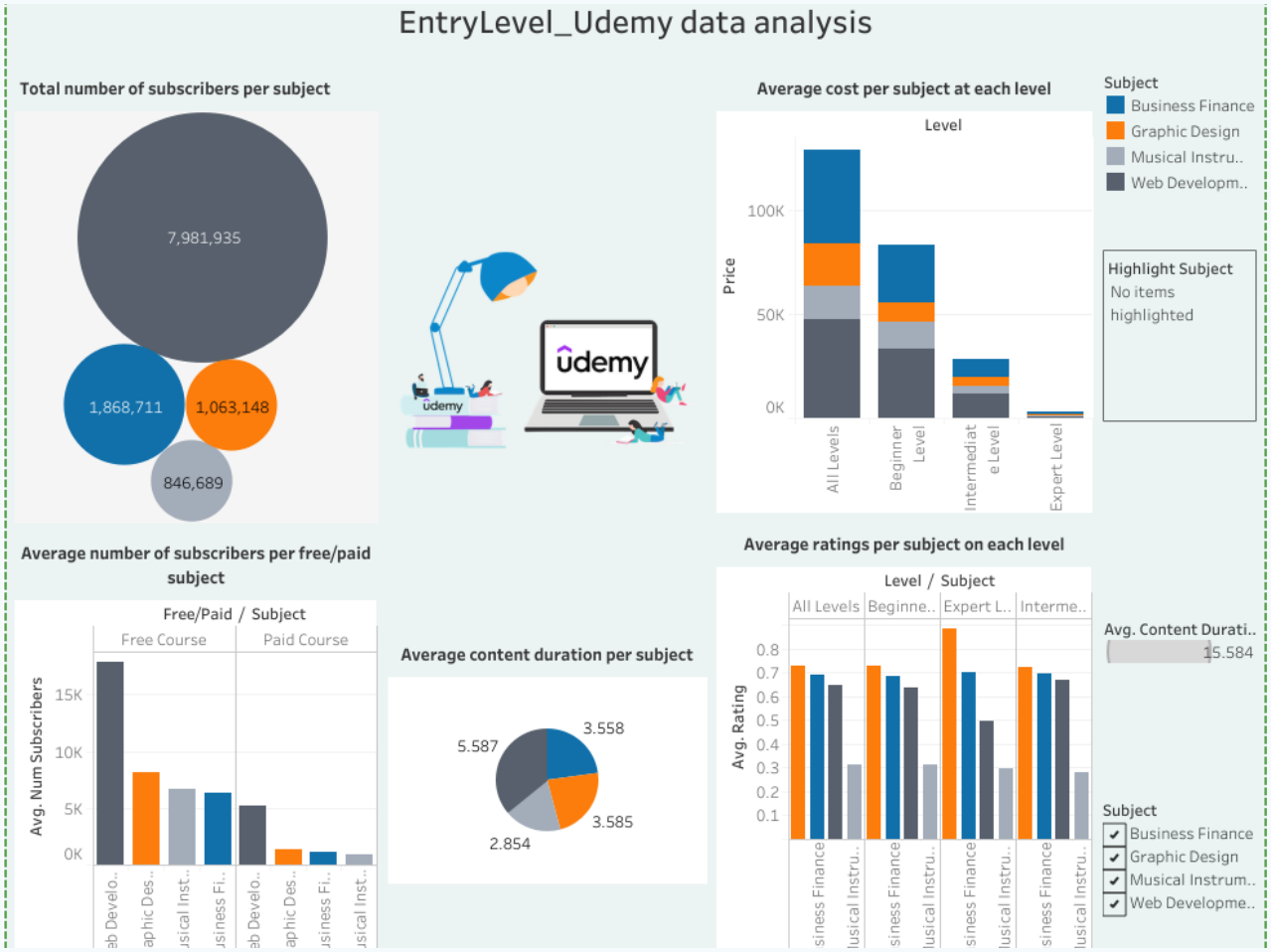
subject	AVERAGE of num_subscribers
Business Finance	1569.026868
Graphic Design	1766.026578
Musical Instruments	1245.130882
Web Development	6635.024938
Grand Total	3199.260881

figure 3: average content duration per subject

subject	AVERAGE of content_duration
Business Finance	3.557696613
Graphic Design	3.585465116
Musical Instruments	2.854387255
Web Development	5.586658354
Grand Total	4.096137106

Findings

- Below is a visualization of the Udemy data set set that was cleaned. it has comparison of the total number of subscribers per subject. Average cost per per level, number of subscribers between free and paid course, average content duration per subject and average rating per subject.



Analysis

It is accessed that according to the provided data set from the four main subjects the Graphic design subject is rated the highest meaning most people are more content with how its been administered.

All level and the beginner level for all subjects are what brings in most money for the paid subjects at Udemy. It has come to light that with the high number of subscribers majority is of those doing the courses being offered for free., which is highly from subscribers doing the Web Development subjects.

Conclusion

Udemy is basically offering skills and more of its time to subjects that are not being paid and that may be part of the reason why they are not making more revenue but offering more of their time and resources to users.

They can consider moving some of the free courses to being paid and even focus on getting higher rating on all subjects because if people see high ratings of on subjects they will be easily convinced that its worth the paid money.



Capstone Project Description

For the Capstone project data from the Air BnB New York City for the 2019 was used.

The purpose of this report was to allow cleaning of the data to take place as well as read data and show possible trends found in the data. The objective was to find trends that follow in the Air BnB of that area from different areas that reside within New York City and to be able to pin piont areas that are performing better compred to the other area as well as to come up with possible solutions based on the given statistics of 2019.

The methods used was consolidating the dataset into one sheet using Google sheets. I then had to clean the consolidated data to avoid working with repeating data. I then used tableau public inorder to show the trends followed by the dataset.

From the data I then analysed that only a certain group in the neighbourhood groups is performing better as compared to the other areas and one is not performing that well. The main reason of that is because of the offering of a certain type of room and having more ratings.

It would be good if the other areas would follow some of the trends being made by other neighbourhood groups.

Data Design

The NYC data for 2019 Air BnB has been consolidated and cleaned using Google Sheets.

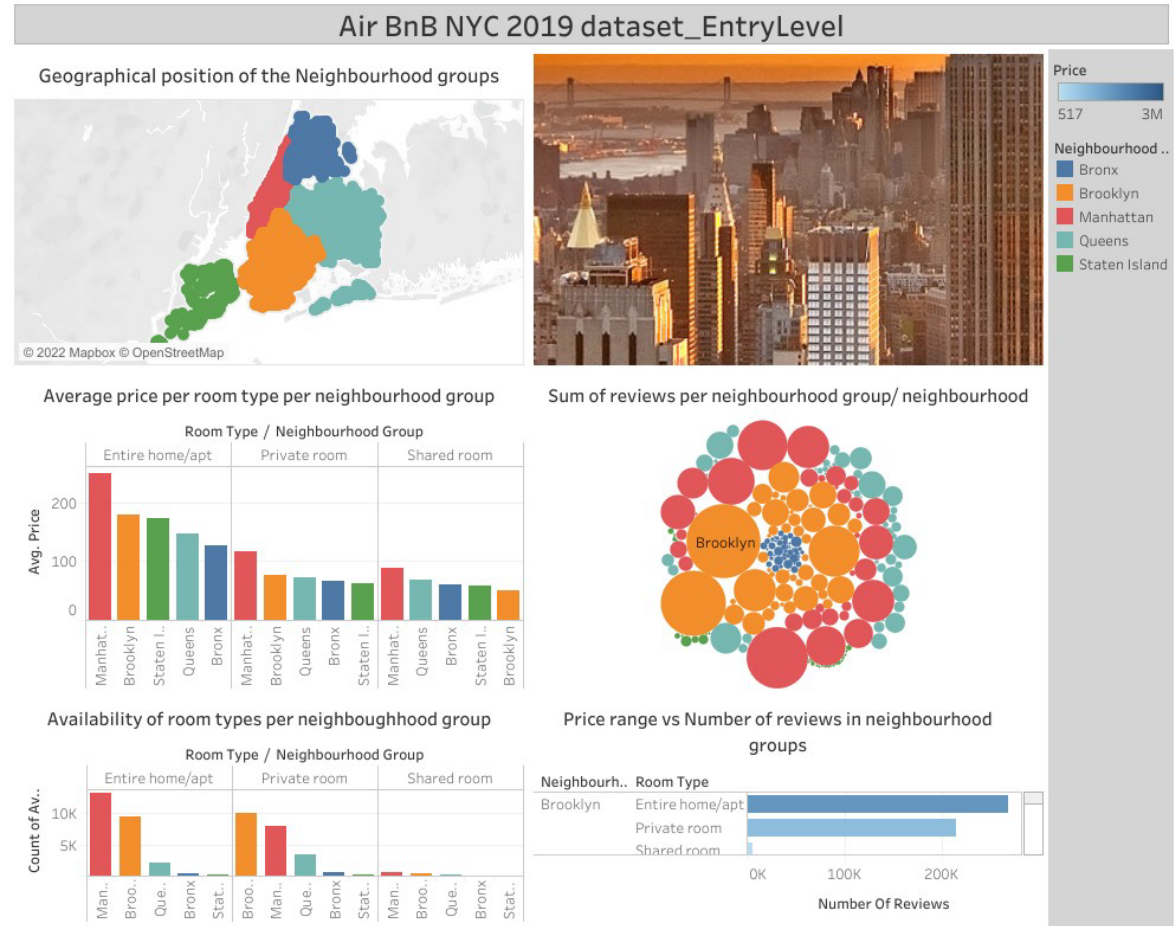
Tableau was used to virtualize the data according to geograhpics, ratings , prices and more. Tableau has been used to represent the virtualisation of the data as it made things easier for looking at trends made by the data .

Findings

- Below is a visualization of the Air BnB New York City data set that was cleaned. It has the geographical positions of the different groups , average price per room, sum or reviews , availability of rooms as well as comparison of the ratings and price.

Figure 3: Sum of reviews per month per room type

room_type	SUM of reviews_per_month
Entire home/apt	26565.34
Private room	25529.62
Shared room	1245.08
Grand Total	53340.04



Analysis

Based on the number of Air Bnb's located around New York City, they have been grouped into 5 groups based on main areas in NYC.

Manhattan has the most rated and yet highest prices amongst other groups making it the top performer in all the groups followed by Brooklyn.

Amongst the room types the best performing one is the Entire house, making it one with most reviews and high price.

Conclusion

Air BnB in NYC during 2019 had the best performance in certain locations with best performers as those housing entire house or private rooms.

It seems having a steep price isn't a problem for most consumers but space that is enough to host their family is.

So for the poor performers they could consider giving their target market what they want, which at the moment is enough rooms to either host their families or private rooms.

Appendix

Google Sheets Data Set for Udemy Project

https://docs.google.com/spreadsheets/d/1y1kILplTbLseRSoKWoj8ba-GEUcqn5i0q7J4FIWAjKQ/edit?usp=sharing

Tableau Data Visualization For Udemy Project

https://public.tableau.com/views/Data_cK_EntryLevel/Dashboard2?:language=en-GB&:display_count=n&:origin=viz_share_link

Google Sheets Data Set for [chosen] Project

https://docs.google.com/spreadsheets/d/1YTR6qiI_S6UPw7N-V_Ytli9bCO2ojiOw1qT5Aiqp3nI/edit?usp=sharing

Tableau Data Visualization For [chosen] Project

https://public.tableau.com/views/AirBnBNYC2019_EntryLevel/Dashboard1?:language=en-GB&:display_count=n&:origin=viz_share_link