Agenda Case Study

agoda

Agenda

- Problem Statement
- Data Overview
- Approach
- Commission Rate Estimation
- Customer Segmentation.
  - Spend Based Classification
  - Customer Classification - Frequency of Booking
- Merging Classifications
- Final Classification

agoda

Problem Statement

- Agoda VIP Program
- How to segregate existing customers.

	VIP Bronze	VIP Silver	VIP Gold	VIP Platinum
Best Price Guarantee	✓	✓	✓	✓
Insider deals	✓	✓	✓	✓
VIP deals up to 12% off	✗	✓	✓	✓
VIP deals up to 18% off	✗	✗	✓	✓
VIP deals up to 25% off	✗	✗	✗	✓
Free breakfast and other perks on selected properties!	✗	✗	✗	✓

agoda

Data Overview

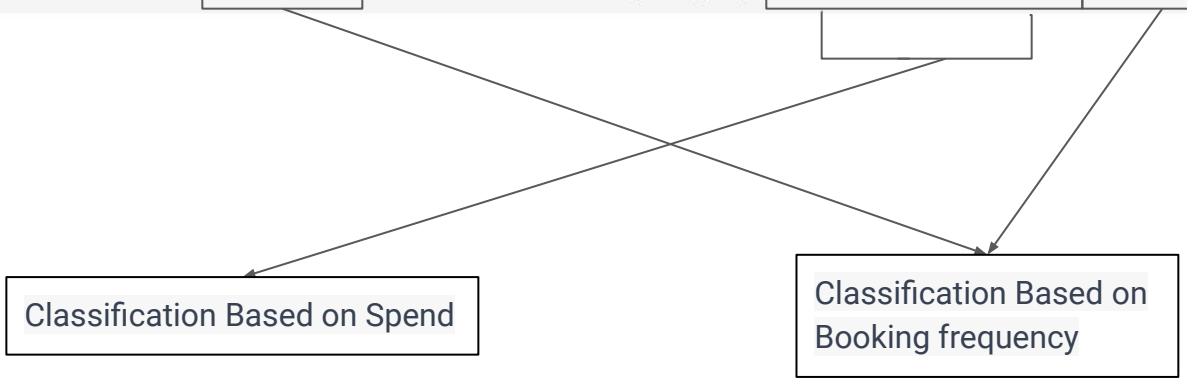
	travelCode	userCode	name	place	days	price	total	date
0	0	0	Hotel A	Florianopolis (SC)	4	313.02	1252.08	09/26/2019
1	2	0	Hotel K	Salvador (BH)	2	263.41	526.82	10/10/2019
2	7	0	Hotel K	Salvador (BH)	3	263.41	790.23	11/14/2019
3	11	0	Hotel K	Salvador (BH)	4	263.41	1053.64	12/12/2019
4	13	0	Hotel A	Florianopolis (SC)	1	313.02	313.02	12/26/2019

40552 rows × 8 columns

Kaggle Data set: https://www.kaggle.com/datasets/atulmittal199174/dataset-of-agoda-hotels

Approach

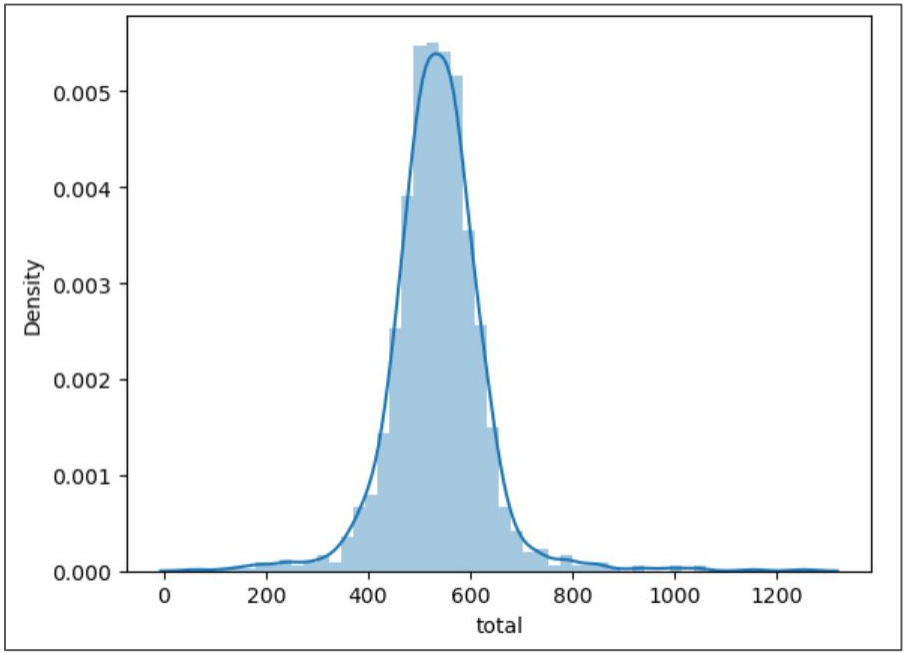
	travelCode	userCode	name	place	days	price	total	date
0	0	0	Hotel A	Florianopolis (SC)	4	313.02	1252.08	09/26/2019
1	2	0	Hotel K	Salvador (BH)	2	263.41	526.82	10/10/2019
2	7	0	Hotel K	Salvador (BH)	3	263.41	790.23	11/14/2019
3	11	0	Hotel K	Salvador (BH)	4	263.41	1053.64	12/12/2019
4	13	0	Hotel A	Florianopolis (SC)	1	313.02	313.02	12/26/2019



agoda

Data Overview

- Group By “userCode” column.
- Plot pdf on total again.
- Total Unique Users: 1310



	userCode	days	total
0	0	2.629630	593.145185
1	1	1.000000	193.360000
2	2	2.694444	558.003333
3	3	2.300000	499.568500
4	4	2.571429	565.316071

agoda

Commission Rate Estimation

- The code defines the range of commission rates as 15% to 22%.
- It performs linear regression using the 'total' column as the independent variable and the 'price' column as a proxy for the commission rate.
- The predicted commission rates are normalized between the defined range and added as a new column ('commission_rate') in the 'hotels' DataFrame.

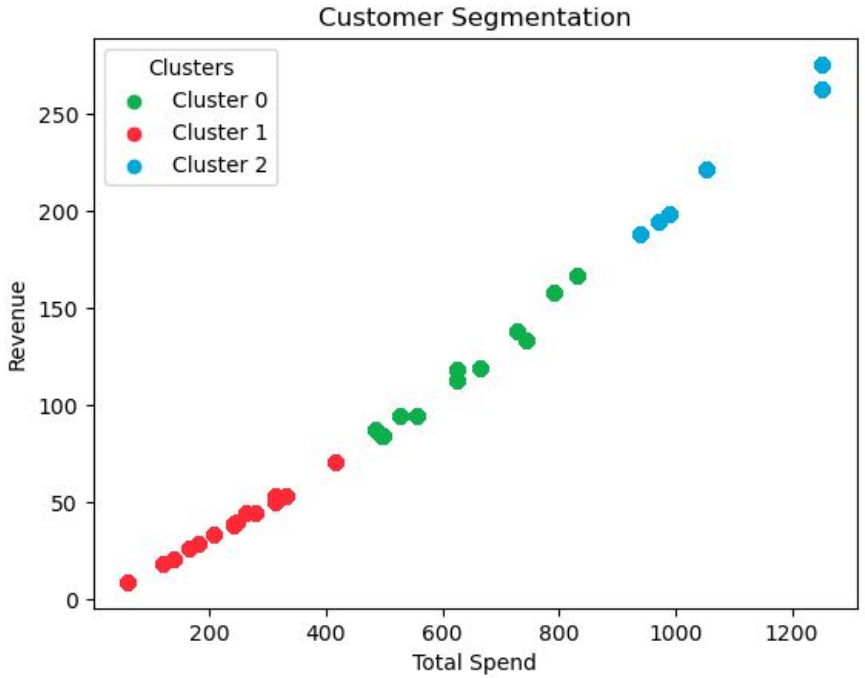
userCode	price	total	commission_rate	revenue
0	313.02	1252.08	0.22	275.4576
0	263.41	526.82	0.18	94.8276
0	263.41	790.23	0.19	150.1437
0	263.41	1053.64	0.21	221.2644
0	313.02	313.02	0.16	50.0832

price = m * total + C

agoda

1. Spend Based Classification

- K-means classification with K = 3



Customer Segmentation - K-means

- Ticket classification: Categorizing customers based on revenue
- Categories: Low Ticket, Medium Ticket, High Ticket

	userCode	revenue	cluster	Ticket_Classification	Preference
0	0	262.9368	2	High Ticket	High Ticket
1	0	94.8276	0	Medium Ticket	High Ticket
2	0	158.0460	0	Medium Ticket	High Ticket
3	0	221.2644	2	High Ticket	High Ticket
4	0	50.0832	1	Low Ticket	High Ticket

2. Customer Classification - Frequency of Booking

- Groups by 'userCode' and 'year', counting the number of occurrences for each combination.
- It then transforms the grouped data into a wide format using the 'unstack' method, filling missing values with zero.
- Categories: Regular, Disconnected, Potential Regular

userCode	2019	2020	2021	2022	2023	Classification
0	5	18	4	0	0	Disconnected
1	2	0	0	0	0	Disconnected
2	4	15	12	5	0	Potential Regular
3	4	16	14	12	14	Regular
4	3	14	13	18	8	Regular

agoda

Merging Classifications

userCode	Preference
0	High Ticket
0	High Ticket
0	High Ticket
0	High Ticket
0	High Ticket

userCode	Classification
0	Disconnected
1	Disconnected
2	Potential Regular
3	Regular
4	Regular

agoda

Merging Classifications

userCode	place	days	price	total	revenue	date	cluster	Preference	Classification
0	Florianopolis (SC)	4	313	1252.08	275.4576	2019-09-26	2	High Ticket	Disconnected
0	Salvador (BH)	2	263	526.82	94.8276	2019-10-10	0	High Ticket	Disconnected
0	Salvador (BH)	3	263	790.23	150.1437	2019-11-14	0	High Ticket	Disconnected
0	Salvador (BH)	4	263	1053.64	221.2644	2019-12-12	2	High Ticket	Disconnected
0	Florianopolis (SC)	1	313	313.02	50.0832	2019-12-26	1	High Ticket	Disconnected

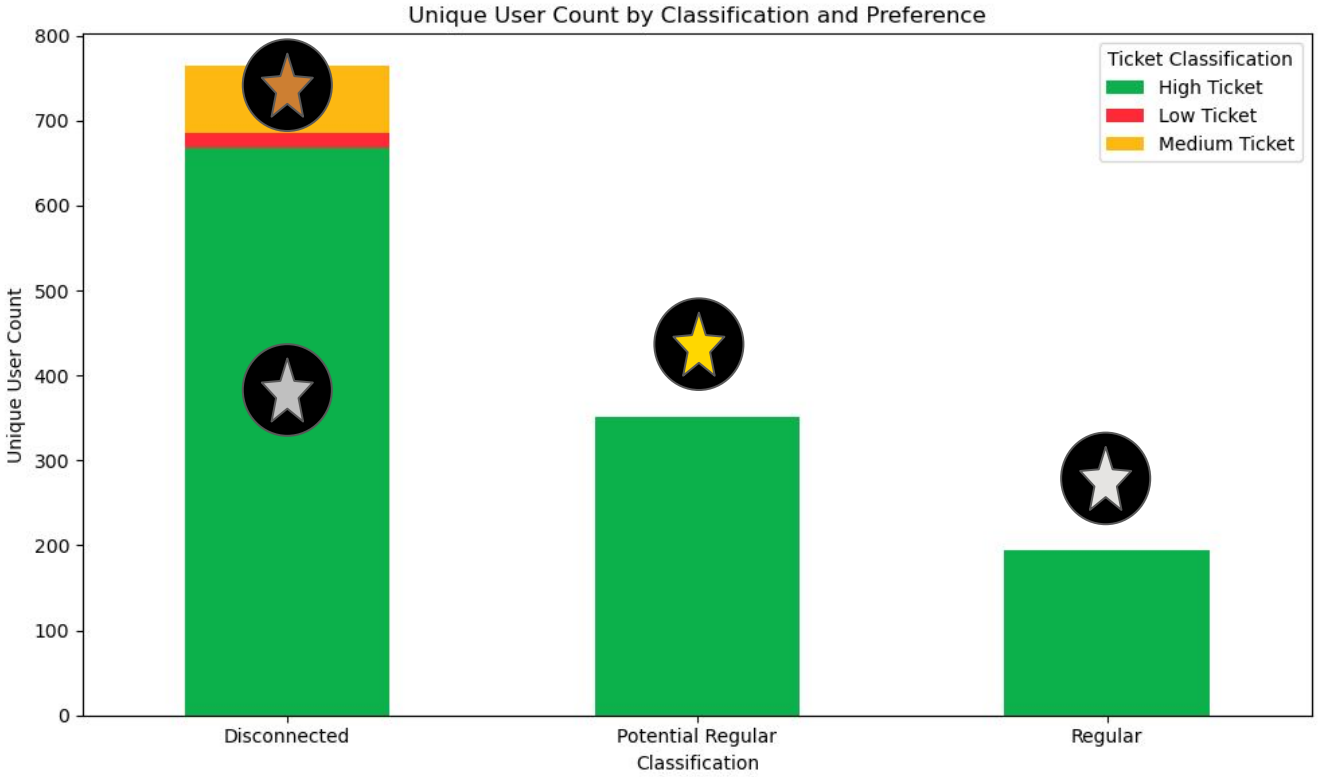
40552 rows

Final Classification

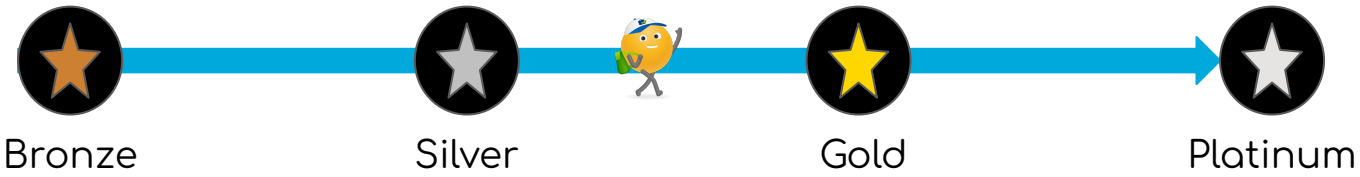
- The code groups the data by "Classification" and "Preference" columns, then calculates the count of unique "userCode" values for each group, resulting in the final classification of users based on their preferences.

	Classification	Preference	userCode
0	Disconnected	High Ticket	668
1	Disconnected	Low Ticket	17
2	Disconnected	Medium Ticket	79
3	Potential Regular	High Ticket	352
4	Regular	High Ticket	194

1310


Unique User Count by Classification and Preference
Ticket Classification
High Ticket
Low Ticket
Medium Ticket
Unique User Count
800
700
600
500
400
300
200
100
0
Disconnected
Potential Regular
Regular
Classification

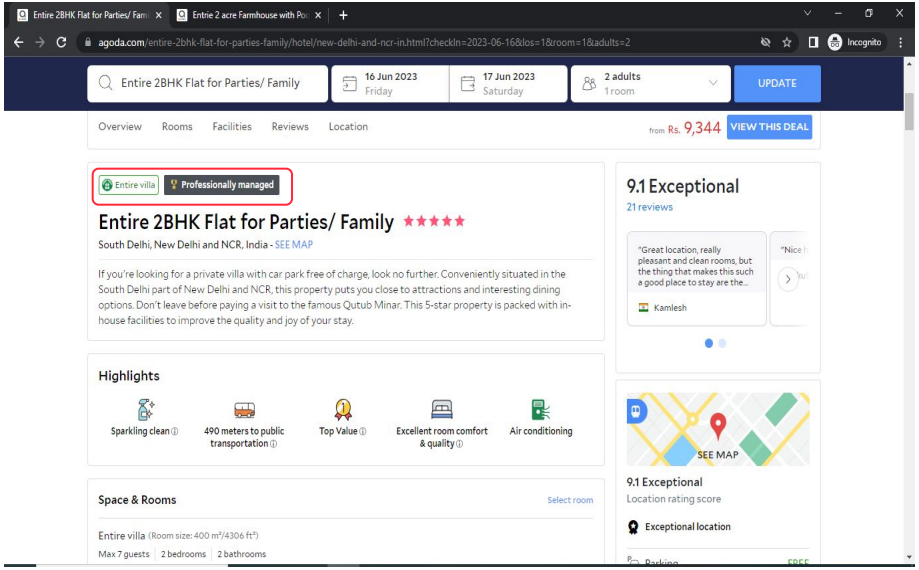
agoda

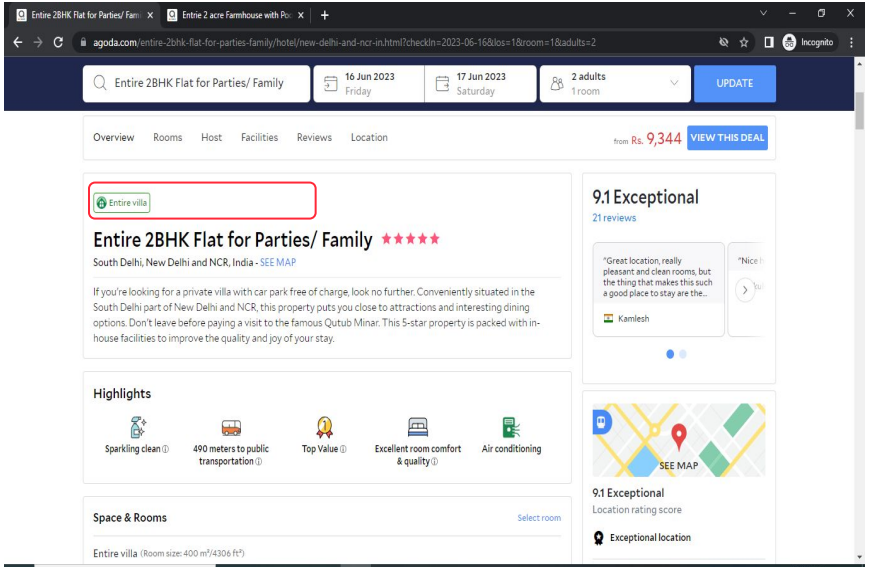


	VIP Bronze	VIP Silver	VIP Gold	VIP Platinum
Best Price Guarantee	✓	✓	✓	✓
Insider deals	✓	✓	✓	✓
VIP deals up to 12% off	✗	✓	✓	✓
VIP deals up to 18% off	✗	✗	✓	✓
VIP deals up to 25% off	✗	✗	✗	✓
Free breakfast and other perks on selected properties!	✗	✗	✗	✓

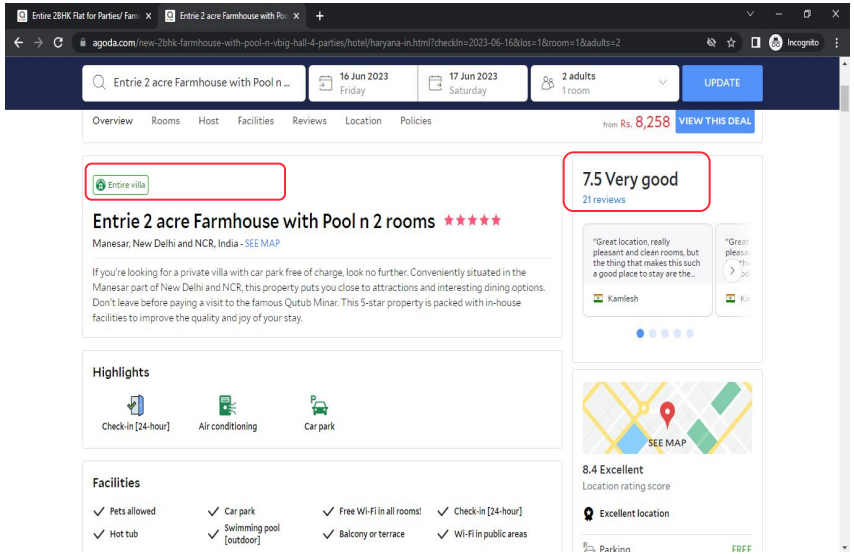
agoda

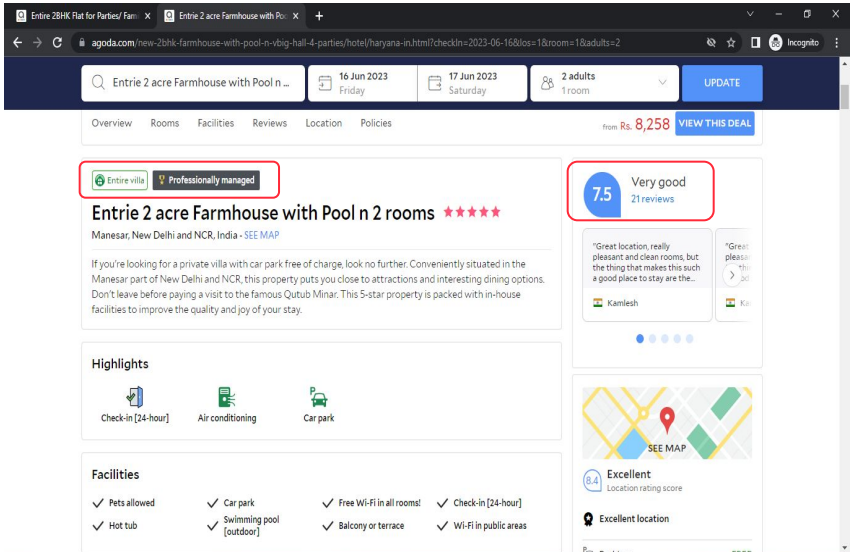
Observations





Observations





agoda