Capstone Project Airbnb

Stephen Fok

Erasmus

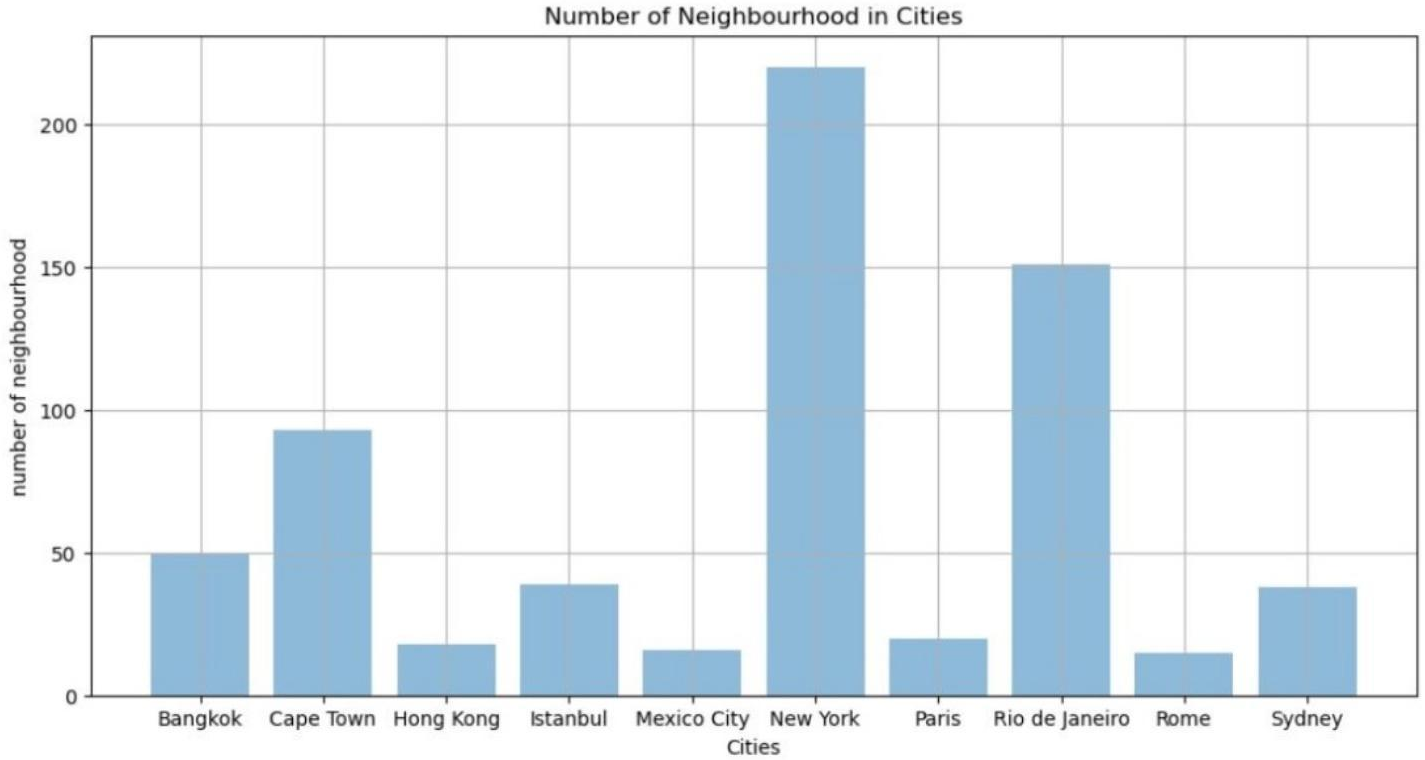
Ka Ho Li

Project Aim & Business Value

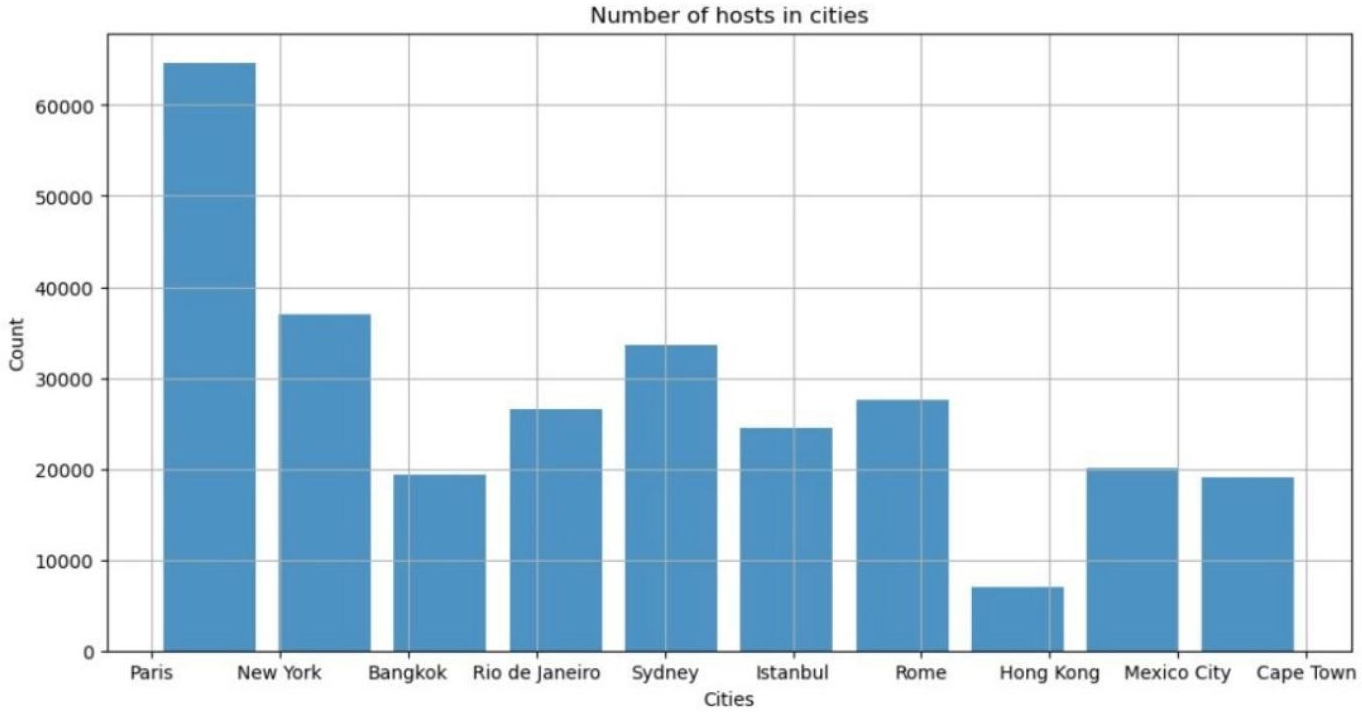
1. Analyze **price** and **occupancy** trend for Airbnb Hong in Kong area

2. Provide a **price range** for users to have budget planning before trip

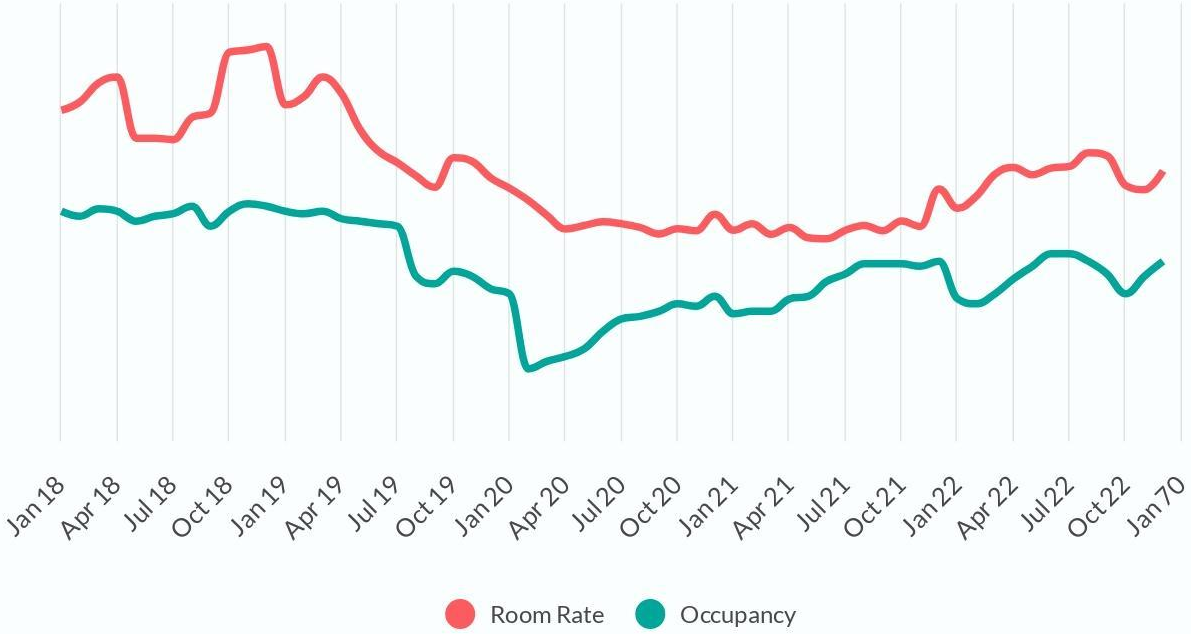
Global Market



Global Market



HK Market Trending (2018-2022)

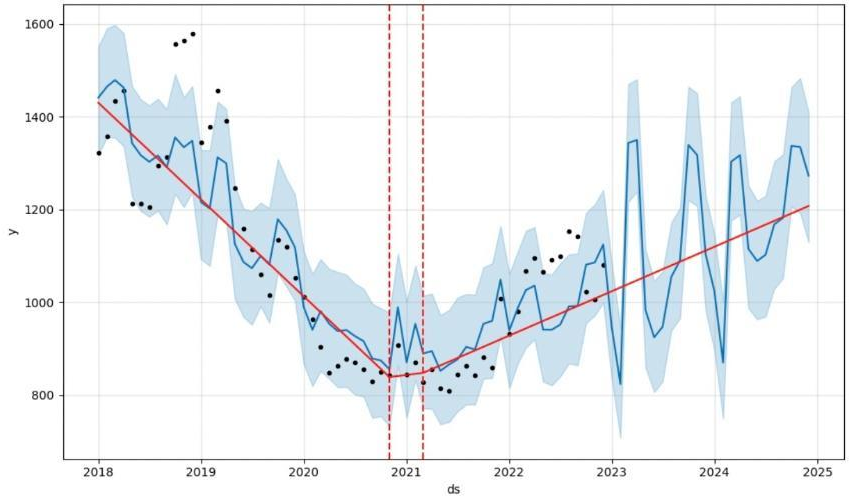


Comparison

- in the middle of 2018, room rate **has dropped** even bore COVID outbreak while we do not see the same pattern on occupacy%
- occupacy% hit its lowest point at early 2020, after that there is **constant increase trend** while room rate still struggle until late 2021

Prediction for Room Rate

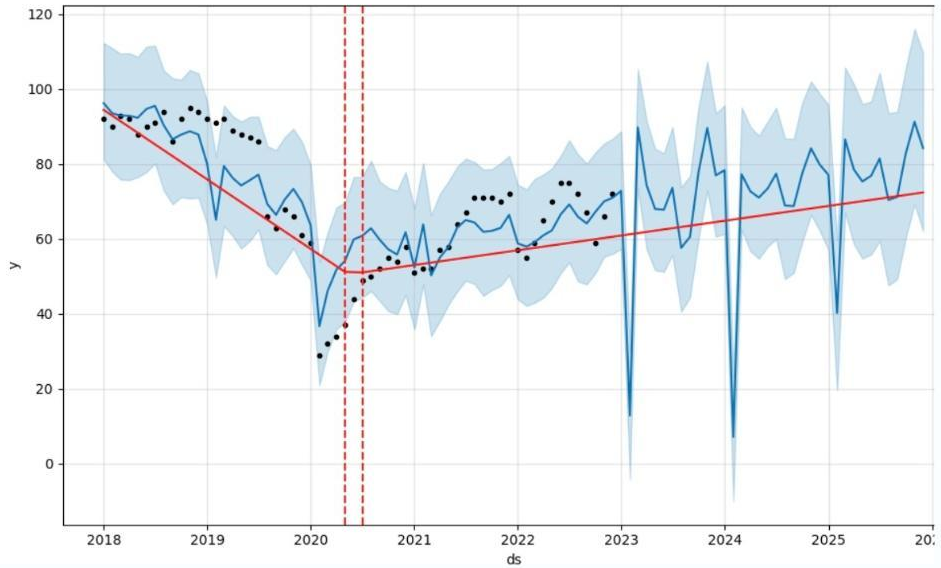
```
In [43]: from prophet.plot import add_changepoints_to_plot
         fig = my_model.plot(forecast)
         a = add_changepoints_to_plot(fig.gca(), my_model, forecast)
```



- Since 2018, it continues its **decreasing trend** and hits the bottom in 2021
- **Red dot line** indicates where changepoints take place, which is also located in 2021
- Mostly, there is an significant **increasing trend** right after 2023

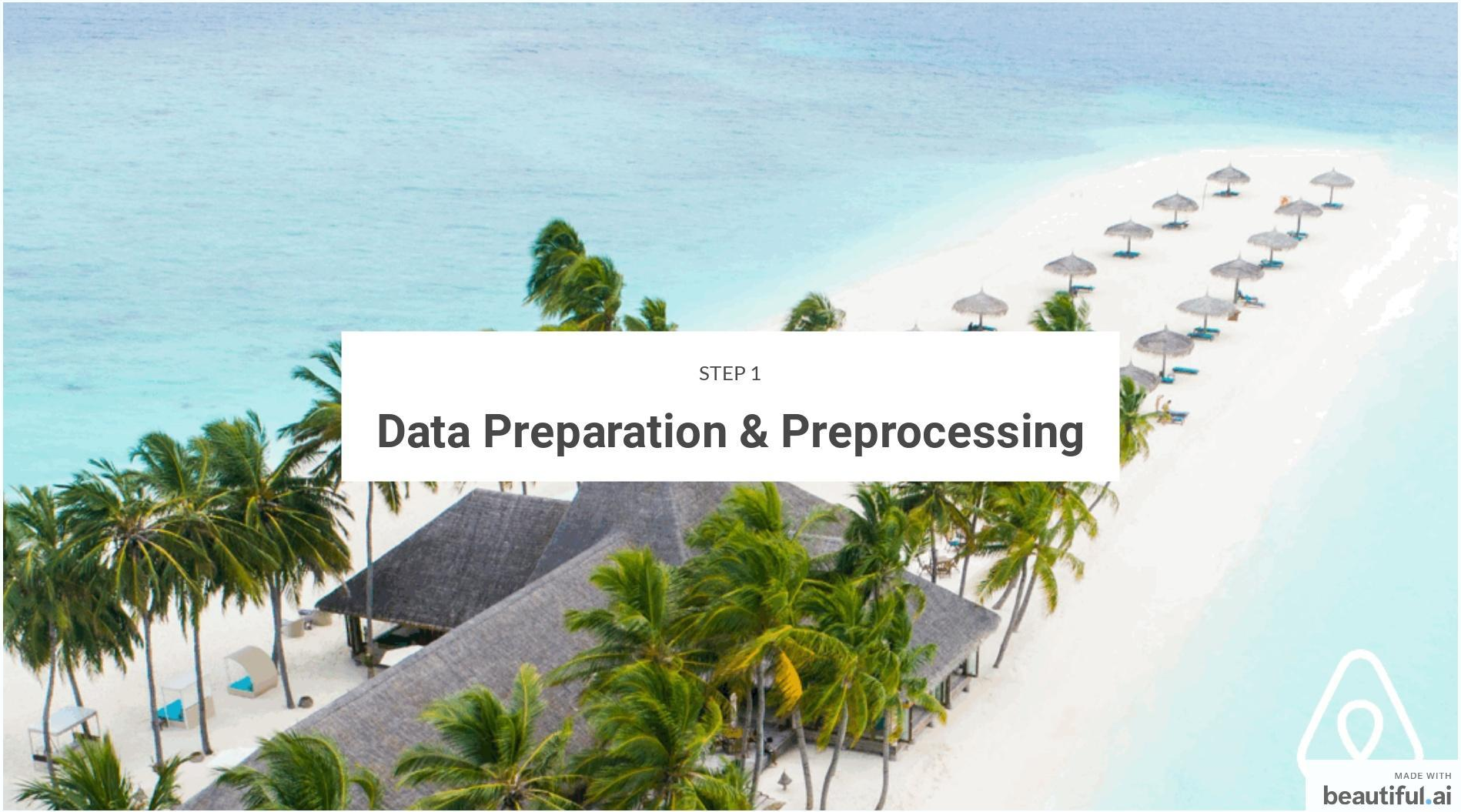
Predicition for Occupancy

```
]: from prophet.plot import add_changepoints_to_plot
   fig = my_model.plot(forecast)
   a = add_changepoints_to_plot(fig.gca(), my_model, forecast)
```



- Since 2018, it continues its **decreasing trend** and hits the bottom in early 2020
- **Red dot line** indicates where changepoints take place, which is also located in 2020
- Mostly, compared to room rate, although there is a increasing trend, the curve is **flatter and more stable**

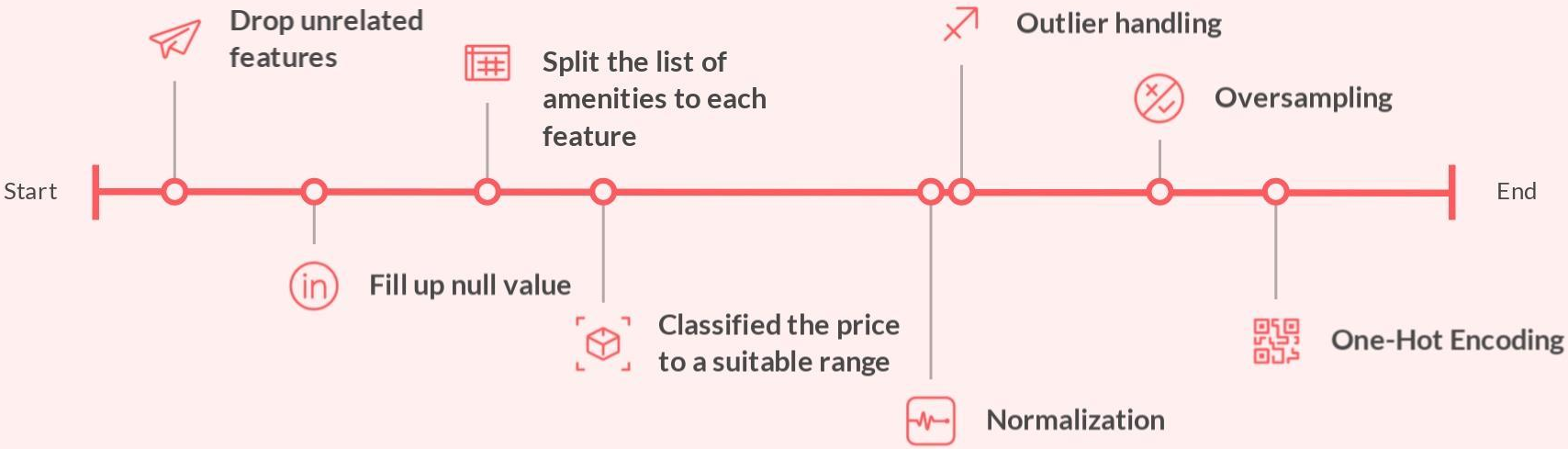
Project Flow



STEP 1

Data Preparation & Preprocessing

Data Preparation & Preprocessing



STEP 2

Data Modelling

Model 1

Algorithm

RandomForest Classification

DecisionTree Classfication

ExtraTrees Classification

GradientBoosting Classification

XGBoost Classification

KNN Classification

HPO

- Giving the weight for each price class
- Tuning the estimators number, Final pick “90”

Outcome

Random Forest Model Accuracy 84.27%

	XGB	KNeighbors	DecisionTree	RandomForest	ExtraTrees	GradientBoosting	GaussianN	LogisticRegressio	SV
0	80.947096	63.891972	82.352941	85.053644	84.831669	59.563448	22.715501	45.874954	61.746208
1	51.655629	37.615894	45.033113	50.993377	51.390728	47.814570	11.125828	34.701987	22.384106

	XGB	KNeighbors	DecisionTree	RandomForest	ExtraTrees	GradientBoosting	GaussianN	LogisticRegressio	SV
0	80.947096	63.891972	82.389937	84.424713	84.646689	59.600444	22.715501	45.874954	61.746208
1	51.655629	37.615894	44.238411	53.245033	51.920530	47.417219	11.125828	34.701987	22.384106

	XGB	KNeighbors	DecisionTree	RandomForest	ExtraTrees	GradientBoosting	GaussianN	LogisticRegressio	SV
0	80.947096	63.891972	82.130966	84.646689	84.609693	59.563448	22.715501	45.874954	61.746208
1	51.655629	37.615894	44.503311	52.582781	50.728477	47.152318	11.125828	34.701987	22.384106

Model(s) Creation, Evaluation and Comparison

Random Forest - HPO

- Giving the weight for each price class

Dataset	1	2	3	Average
1) Drop 7+7	85.534591	85.386607	84.794673	85.238624
2) Drop 7	57.218543	56.953642	56.821192	56.997792

- Tuning the estimators number, Final pick “90”

	acc1	acc2	acc3	avg
est				
90	85.312616	85.423603	85.793563	85.509927
100	85.460599	85.645579	85.238624	85.448267
98	85.423603	85.423603	85.386607	85.411271

Model 2

Algorithm

RandomForest Regressor

It has many parameters that can be adjusted to optimize the model, and is generally accurate and powerful, even with non-linear relationships between features

Features Selected

- Number of guests to accommodate
- Property Type
- Number of Bedrooms
- Location (Neighbourhood)
- Instant Bookable
- Number of Amenity

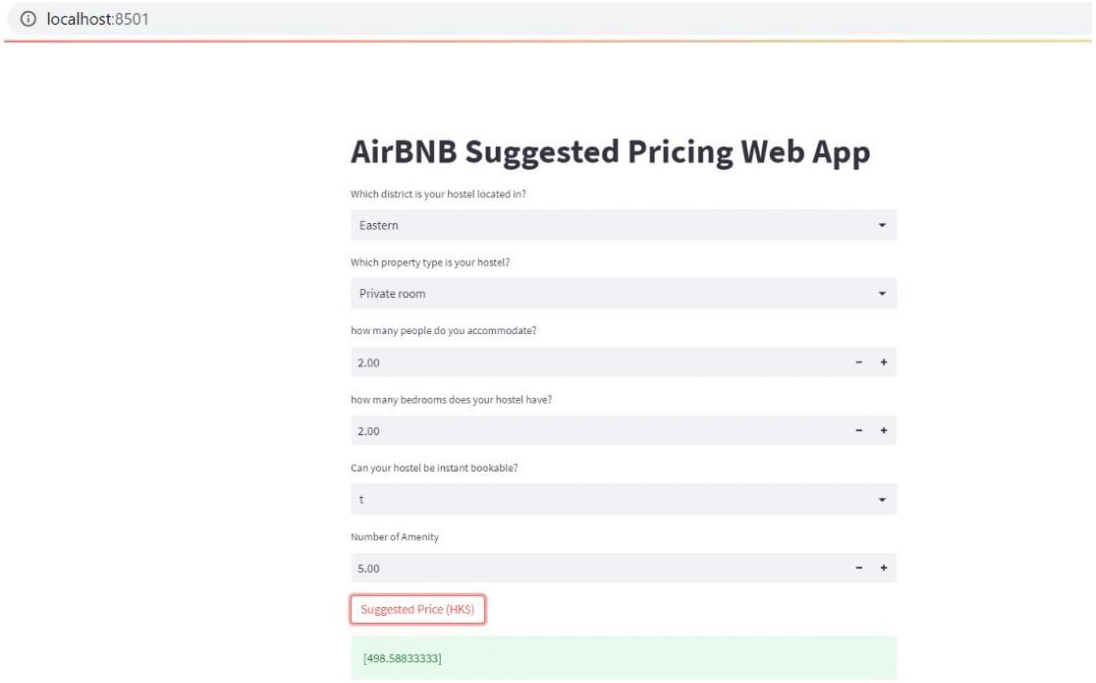
Outcome

Mean Absolute Error 218.9

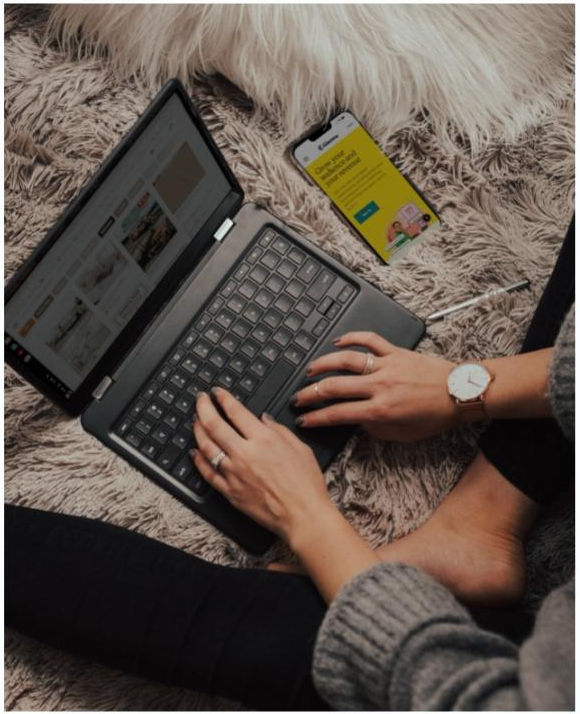
STEP 3

Model Deployment

Steamlit Web App



Limitations and Challenges



1. extreme figures lead to volatile prediction
2. null data and outliers in the dataset
3. set price range in order to get higher accuracy during modelling
4. fitting scaling data into streamlit

World Tourism Organization Review & Forecast

tourist arrivals around the world could reach in 2023

80-95% of pre-pandemic levels

best performed region during 2022

Europe

world-wide tourists travelled in 2022

900 million

Conclusions

- The travel industry is showing signs of increasing optimism in 2023.
- Specifically, the Asia-Pacific region is expected to see an improvement in 2023, with the region's growth and the recover of China.
- All these are very encouraging factors to support both room rate and occupancy for Airbnb so it is quite clear that Airbnb will benefit from it.
- Although there may still need some time for the whole travel industry to fully recover, we may not have to wait that long.

Thanks!