

Speaker Notes

Slide 1:

Hello everyone. Today we will be presenting on Airbnb Toronto data and its Price Prediction.

Slide 2:

- As you all may know, Airbnb is a popular online marketplace for short-term rentals. Airbnb hosts can rent their entire house/apartment/private or shared rooms to guests from around the world for a certain price.

Slide 3:

- Although Airbnb provides the host with general tips regarding the default nightly prices of their properties upon listing, the final decision regarding the property price ultimately rests with the hosts. There are many factors to consider while determining the price from location, to the room type to the amenities offered which can sometimes make it a challenge to come up with the right price for their listings. So the business problem we are looking to solve is “How can hosts successfully determine the optimal rent prices of their properties?”

Slide 4:

- Our aim is to provide the hosts with a data-driven solution through exploratory analysis and machine learning that can predict the rental prices for each property.
- So we constructed a data driven solution that uses regression based machine learning models to predict the prices for these Airbnb listings. Some of the features we have used in our machine learning model are location, different property types, amenities, customer ratings and reviews, etc.

Slide 5:

- Here we have shared some important insights that we could find from our data through the Exploratory Data Analysis phase. The first chart on the top shows the majority property types, from this we see that most of the properties in Toronto are Apartments this is especially true in Downtown Toronto.
- 1 room is popular in Scarborough and North York apart from Downtown. Entire home category is quite limited compared to the others. The chart below shows the average prices for these property types in different boroughs and we can see that renting the entire house can get the highest price - especially true in Central Toronto and Downtown Toronto. Apartments are relatively cheaper compared to the houses in all boroughs.

Slide 6:

- We can see the average price for the property by the number of people accommodate. The properties that can accommodate single guests have a low price of approx. CAD \$64 whereas properties accommodating 14 people can have an average price of CAD \$570 per night.
- After the number of guests accommodated exceeds 4, the prices go above CAD \$200
- Internet is the most popular in all listings, hosts can charge higher prices for providing other amenities like home appliances, high end electronics, coffee machines, parking

Slide 7:

- Here is a demo of an interactive feature from our dashboard. This interactive map allows the end user to learn about the average prices in each borough. The user can adjust the price range as well as the boroughs they specifically want to see and derive more meaningful and personalized insights.

Slide 8:

- I'll now move on to our Machine Learning model. Since we were predicting prices we needed a regression model. We started with Linear Regression, but the results were not accurate. So, we tried several others.
- Our metrics were RMSE, which determines the difference between predicted and real values, and R squared, which shows how much of the price can be explained by the features in the model. This table shows how each one of the models performed compared to linear regression.
- As you can see, XGBoost Regressor gave us the best results, with an RMSE 14% better than linear regression. We tried to further improve it by using `gridSearchCV` to tune the hyperparameters. Resulting in an RMSE 15% better than our first model.
- On the right side, the most important features are listed, location, property type, and accommodation are the most important as we mentioned earlier in our data analysis.

Slide 9:

- Here is a demo of our application, where the host can input information about their profile and their properties. The application, then, predicts the optimal price for listing on Airbnb.

Slide 10:

- The next steps should include more feature engineering, since, right now, our model is predicting prices significantly higher than the real values.
- Some data was left out from this delivery, like seasonality. Although we have the data, we decided to keep it for further improvements. As well as performing sentiment analysis on customer reviews to understand if they have any impact on pricing.
- I also would like to mention the limitation of our dataset. Prices can be affected by many factors that were not available on the Inside Airbnb data we have been working with, such as demand, points of interest nearby, payment options, etc.

Slide 11:

- Check out our project's github page for more details on our project from our machine learning model, code files and dashboard.

Slide 12:

- Thank you for listening and we are now opening the floor for any questions