**PREDICTION AND RECOMMENDATION**

**ENGINE FOR AIRBNB**

**CAPE TOWN SOUTH AFRICA**

CRISP-DM Term Project Report

**TEAM CONTRIBUTORS**

Mercy Onduso

Nurulain Abdi

Amos Kibet

Beth Mithamo

**POWERED BY**

MORINGA SCHOOL

**INTRODUCTION**

**MOTIVATION:**

Machine Learning and Artificial Intelligence are at the forefront of innovation in today’s world. With Data being the most valuable resource in the 21st century more and more data-driven business decisions are being made with machine learning and large-scale data analytics.

Over the last two decades, hotel industries have seen the power of their brands’ promise move away from their marketing teams into the hands of the consumer. Aided by the rise of online review platforms like TripAdvisor, hotel prices can easily be compared with other properties in the market. Considering this competitive market, every hospitality brand is trying to provide better services to customers by utilizing machine learning advances in the technology sector. Hence, in this project, we proposed a few competitive solutions useful to Airbnb users and hosts. We are using huge data of Airbnb Cape town South Africa to create a recommendation model to recommend listings to the users using collaborative filtering algorithm and predict optimum listing price to the beneficiaries.

**Overview:**

Airbnb is a global online marketplace that offers housing and other accommodations to travelers. The platform has grown significantly in popularity over the years, with millions of hosts and guests using the platform for their travel needs.

The aim of this project is to create a recommender system that will help stakeholders and clients have a better strategy in decision making. The system will also help stakeholders do proper renovations of their listings efficiently without inconveniencing their clients.

**Business Problem**

Airbnb has become a popular alternative to traditional hotels for tourists and visitors in Cape Town, South Africa. However, despite its many advantages, users often face several challenges when using the platform. These include poor recommendations, unreliable pricing, and subpar customer experience. Moreover, stakeholders often struggle to renovate their listings to meet the needs of their target customers. A South-Africa based housing company wants to venture into the Airbnb business and needs to create a sustainable and profitable business model that can compete with established players in the market. The company's stakeholders aim at ensuring customer retention, customer satisfaction and boost their business as a new party entity in the Airbnb Platform. As Data Scientists, we are expected to address questions as well as provide recommendations.

Some of the questions we are expected to answer are:

1. What is the best month to visit Cape Town if you are on a budget?
2. What is the best time to list your property on Airbnb? And how do set price rates according to the time of the year?
3. What is the best time in the year when owners can take down their listing for maintenance and repair?
4. When is the best time to lure clients with offers in the case of an upcoming low season: Time series analysis.
5. **Data understanding**

**Data Source**

We extracted the data from InsideAirbnb which has data from the Airbnb platform. The link to the dataset is provided here: <http://insideairbnb.com/get-the-data/>

The data from the Airbnb app provides insights into the availability, pricing, and characteristics of short-term rental properties, such as apartments, houses, and rooms. The data can be used to understand the demand and supply dynamics in the market, as well as the preferences of guests and hosts. The data can also help identify trends and patterns in guest behavior, such as popular locations, amenities, and property types. Additionally, the data will be used in the development of a recommender systems that can make personalized recommendations to guests based on their preferences and past behavior. This will help both stakeholders and their clients have better strategies during decision making.

**Technologies & Tools used**: Libraries: pandas, numpy, matplotlib, Surprise, Spark MLib, NLTK, scikit Python 3 with Jupyter Notebook, GitHub for repository and collaboration.

**Experiments / Proof of concept evaluation Dataset(s) Used Dataset**

Link to source code: <https://github.com/MercyMoraa/InsideAirbnb>

The dataset includes the following:

1. cale\_df.csv: Details about listing, date, availability, and price in Cape town south Africa.
2. listings\_df.csv: The description of host and location of the property, price, last review, and availability. neighbourhoods.csv: For a particular listing, details of the neighborhood group.
3. reviews\_detail.csv: Review comments by the user for a specific property.

**Model 1** - Recommend Listings using content based filtering contains information like reviewer\_id, listing\_id, and Rating. This data is used in order to predict ratings to listings, which user has not visited/reviewed.

**MODEL 2**: User-Based Collaborative Filtering use review\_scores\_rating. Incorporating knnWithMean

**Model 3**: Recommendation System for Sentiment analysis use reviews\_df together with vader to make analysis and recommender system.

**Model 4:** Time series Analysis.

**Methodology followed:**

**Model 1:**

* Feature selection of the listing\_df.
* Checking for multi-collinratiy.
* Checking for distribution.
* Standard scalar and one hot encoding.
* Checking for similarity.
* Building a recommendation.

**Model 2:**

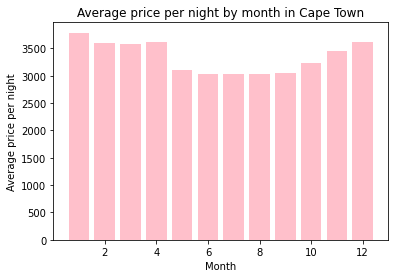
**Model 3: Sentiment Recommender system**

* Loading the data.
* Reviewing the comments.
* Sorting and cleaning the comments.
* Detecting the language.
* Estimating the polarity.
* Build a model base on polarity.

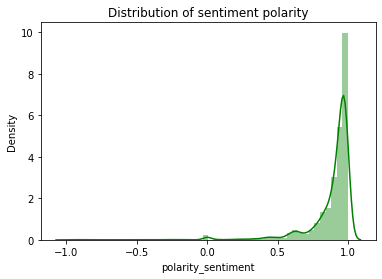
**Model 4: Time Series**

**Graph**

The graph below show the best times to visit Cape Town when the prices are low. The favorable months to go on a trip is July to August.



The next graph shows the sentiment polarity score with VADER. It is skewed to one side which is the positive side which means most places have positive reviews



**Analysis of results**:

**Model 1:** Content based Modelling

Based on our findings, our data is skewed to one side making our data lean on the positive polarity mean we have good hotels. There was no need to include hyperparameters.

**Model 2:**

**Model 3:** Sentiment base recommender system.

The best mode to evaluate was to use RMSE or MSE of the polarity sentiment.

The k value of 150 and the RMSE score is 0.04208087877889592 was the best parameter with the lowest RMSE

**Model 4** :Time series analysis

**Discussion & Conclusions**

**Decisions made:**

1. **Best month to visit Cape Town on a budget:**

This the month of July through August of low budget.

1. **Best time to list your property on Airbnb and set price rates:**

It may be beneficial to list the property during the summer months when the occupancy rate is high and set higher prices during this time.

1. **Recommendation System:**

After comparing different collaborative filtering algorithms, we decided to go for matrix factorization as it gives better RMSE and supports large dataset. SVD is better-provided dataset is not too sparse. Model using SVD cannot be scaled to support large and sparse dataset.

**References**

1. [How Did We Build Book Recommender Systems in an Hour Part 1 — The Fundamentals | by Susan Li | Towards Data Science](https://towardsdatascience.com/how-did-we-build-book-recommender-systems-in-an-hour-the-fundamentals-dfee054f978e)
2. https://www.youtube.com/watch?v=Alu\_cCXNS-k