

## **OPTIMIZING HOST PRICING AT AIRBNB USING BENCHMARKING AND EXTERNAL DATA WITH FEATURE ENGINEERING**

### **INTRODUCTION**

Airbnb has changed travel by making it possible to find a variety of accommodations, catering to all budgets and modes of travel, in competition with the hospitality industry by serving a customer base of 7 million. A balance is needed to achieve this success. Each host lists the properties they own and sets their prices, while guests seek a perfect rental based on their needs. However, with millions of listings and vast user data, optimizing pricing strategies for both hosts and guests can be difficult. Given a single team of data scientists, it is challenging to take all the data science concerns into account as determining the optimal price for any specific rental can be a formidable task for both hosts and guests. In this case study, I propose two approaches i.e., implementing benchmarking existing solutions and harnessing external data sources with feature engineering.

### **PROPOSED SOLUTIONS**

**Solution 1 – Benchmarking with Feature Engineering** – In this solution, Airbnb gets to develop new features by comparing their products by collecting data from other successful rental businesses and the strategies that are used to price their products. With this, we can think of creating new features that will help us in gathering data for our prediction. For instance, one of the ways would be to introduce “guest reviews” where positive reviews and high ratings indicate a desirable listing that delivers a great experience by validating it with locations and facilities as data inputs. We can make sure to get data for this by giving a discount to customers who leave an honest review which helps to predict accuracy and allows us to enable more efficient learning.

Here are the following steps to achieve this solution:

1. **Baseline Model:** Creating a basic model in this step would help to predict future models. In this way, the first model acts as the benchmark against which all other models can be compared to get accurate results.

2. Feature Engineering Teams: The data science team can be segregated into sub-teams focusing on a specific area which helps to concentrate on individual strengths.
3. Benchmark & Sharing Reviews: Each team creates models using their chosen features, and then compares and discusses them with each other, sharing knowledge and seeing which features can be implemented or discarded.
4. Refining Model: Based on the finalized features from each team, an accurate pricing model can be created.
5. Balancing Complexity: They will analyze the degree of complexity needed for an accurate model that can achieve its business goals. By doing this, Airbnb can determine the right balance between model performance and interpretability when it comes to meeting its commercial objectives.

**Solution 2 - Harnessing External Data with Feature Engineering** – A presence of hidden factors is possible that might be influencing the pricing and are out of Airbnb's scope. In such scenarios, it is a good idea to gather more data from external sources. Feature engineering can be employed at this step to make it more compatible with the model. Here are the following steps to achieve this solution:

1. Identify External Data Sources: We can explore external data sources like event calendars, weather forecasts, and even economic reports that might fill those missing gaps. The data science team lead will assess what data to and integrate.
2. Feature Engineering for Integration: This external data might not be in the same semantics as our initial model. We'll need to transform the external data into a format that our model can understand. Essentially, we're giving the model new attributes to understand better.
3. Model Selection & Evaluation: Now, with the final solution, we can decide whether to use two separate models or combine them into a master model (that gives more and better accurate

results). We can then get to weigh the pros and cons to find out if the extra accuracy is worth the added complexity.

This combined approach increases the model's scope for potentially more accurate price recommendations by offering powerful tools for data science leaders in the Airbnb price prediction project. These solutions enable leaders to make data-driven decisions, boost teamwork with group discussions, and weigh trade-offs by balancing model complexity.

**Table - Data-Driven Price Prediction at Airbnb: Project Management Roadmap**

<b>S. No</b>	<b>Steps</b>	<b>Time (days)</b>	<b>Reason</b>
1.	Defining Business Objectives	3	We perform this step to discuss primary goals which can be how to improve Airbnb hosts' booking rates, boost overall revenue, and provide competitive and personalized pricing recommendations.
2.	Preparing and Researching	7	This step helps perform thorough research on how and where data can be gathered and then analyze it to understand its structure, identify issues, and explore how various factors impact pricing. We'll also research competitor models to gain an advantage.
3.	Designing the Solution	14	Our system gives hosts the ability to set prices based on data through creating informative functions; this involves choosing the right machine learning method as well as giving simple instructions.
4.	Aligning Stakeholders	3	This step is important as it helps us to communicate project goals, approaches being used, and timelines ensuring everyone's aligned on expectations and resource needs.

5.	Delivering the Solution	21	In this, since we perform extensive data cleaning, rigorous model training with cross-validation, and continuous improvement via host and guest feedback loops, it ensures our price prediction system remains effective over time.
6.	Measuring Business Impact	3	To measure the system's success, we'll track key metrics (booking rates, revenue, satisfaction) and create reports to communicate its impact on business goals.

## JUSTIFICATION

As the amount of data is increasing, the solutions highlighted above help address the challenges of making data-driven decisions that help Airbnb grow its business. Hosts need a lot of information regarding accurate timely pricing to make better income from the rental listings posted. Benchmarking existing solutions allows Airbnb to learn from its competitors and build upon all the best practices. It helps us learn from industry leaders, fostering collaboration within our data science team and letting us choose the optimal approach. Furthermore, including external data sources widens the scope of data analysis, leading to more informed recommendations. In addition, by identifying the broader variables that influence guest demand, a deeper context can be added than that provided by Airbnb's platform and could lead to more accurate pricing recommendations based on external data like region events or weather trends.

## POSSIBLE CHALLENGES AND MITIGATIONS

While implementing the solutions proposed above, there might be a few challenges we might encounter as a team. They are:

- **Data Availability:** Since we will be wanting to include data from multiple external sources, we must have access to all such relevant data and must be able to address missing values/inconsistencies.

- **Feature Engineering Expertise:** The existing team might not be the SMEs (Subject Matter Experts) on feature engineering which might cause roadblocks while implementing the solutions. As a project manager, I might want to upskill my team on feature engineering techniques and encourage knowledge sharing.
- **Bias in Feature Selection:** Mitigating bias through clear selection criteria based on domain knowledge and statistical significance.
- **Time & Resource Constraints:** Prioritizing feature engineering efforts and utilizing cloud resources when needed. Setting realistic timelines and milestones.
- **Model Interpretability Trade-off:** Balancing model complexity with interpretability could be a potential challenge that can be tackled by documenting the rationale behind choices and using visualization techniques. Considering interpretable models alongside complex algorithms will also be beneficial.

## **CONCLUSION**

To facilitate the optimization of host prices by Airbnb, which employs both benchmarks as well as external information, a data-driven approach should be used. Benchmarking leads to the discovery of good practices, and external information makes it possible to consider market dynamics while making sure that the two kinds of data complement each other and are interpreted in a manner that makes the model workable. This is where the proposed solution works. The hosts become empowered, and the market's competitiveness is improved through these data-oriented strategies.

## **REFERENCES**

1. Concept and solutions for challenges: **How to Lead in Data Science – Book by Jike Chong and Yue Cathy Chang**
2. Benchmarking for Machine Learning:

- a. Machine Learning Mastery: A Gentle Introduction to Benchmarking in Machine Learning:  
<https://machinelearningmastery.com/>(<https://machinelearningmastery.com/>
3. External Data for Airbnb Analysis: <https://insideairbnb.com/>(<https://insideairbnb.com/>