Airbnb Rental Analysis in Barcelona

Explore Barcelona's hidden gems through Airbnb data analysis. Predict rental prices using statistical modeling and identify factors that influence them.

Introduction

Airbnb is an online marketplace that allows people to rent out their homes or apartments to travelers seeking short-term accommodation

Data Source: Data World

Data Period: January 2019 to December 2019

Data Model: Linear Regression, Random Forest and XGBoost

Dependent Variable: Price

Independent Variables:

- Property Type
- Latitude
- Longitude
- Room Type
- Neighborhood
- Number of Reviews
- Cancellation Policy
- Property Type vs Room Type
- Availability
- Maximum Occupancy
- Reviews

Research Questions

- How various factors influence property rental prices in Barcelona?
- Which areas attract majority tourists in terms of neighborhoods?
- Potential audience: New and Existing Hosts
- Price analysis and its dependency on various parameters can help new and existing hosts to competitively price their properties vis-a vis existing rental accommodations.

Initial Data Exploration

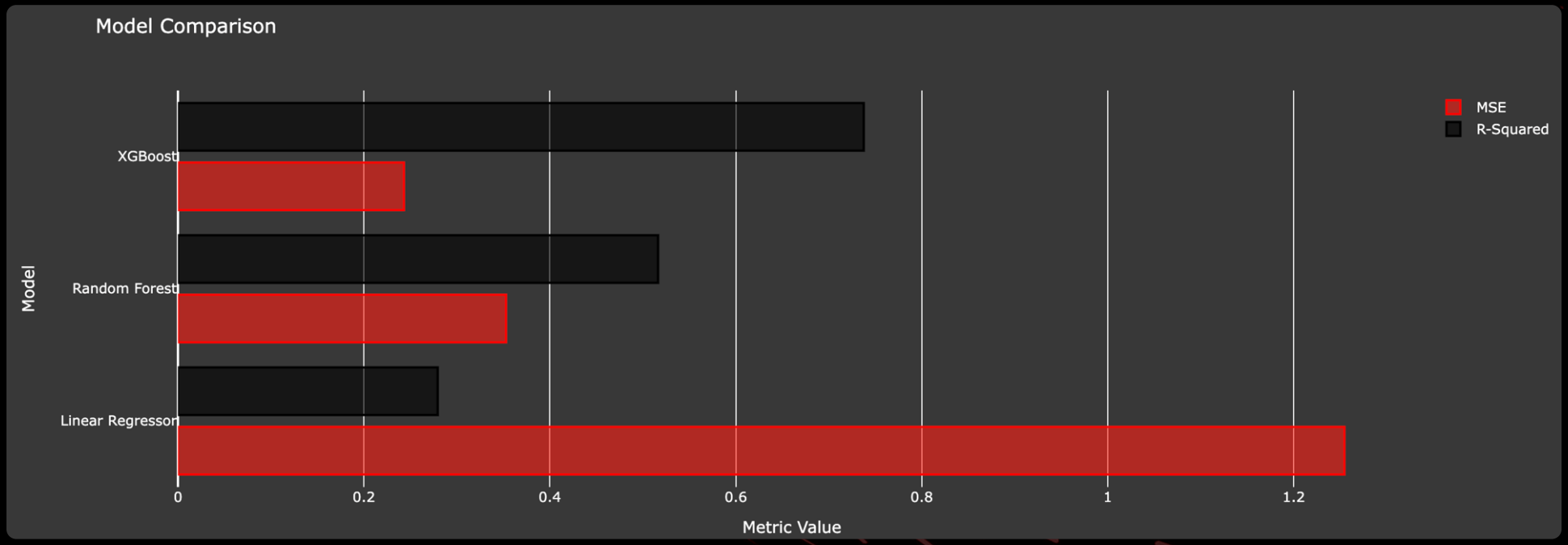
1. Why?
    - Helps identify trends, patterns, and anomalies in the data.
    - Allows for better understanding of the data and its characteristics.
2. How?
    - Consider the business objectives or research questions and select columns that are most relevant to them.
    - Eliminate duplicate or redundant columns that contain the same information or have too many missing values.
    - Iterate the process until the desired number of columns is reached.
3. Uses
    - Developing predictive models.
    - Improving the accuracy and reliability of data used in models.
    - Assessing the quality of data and identifying areas for improvement.

Predictive Modeling

Comparison between Linear Regression, Random Forest and XGBoost:

- Linear Regression is simple and easy to interpret, but may not capture complex relationships.
- Random Forest can handle non-linear relationships and outliers, but can be computationally intensive.
- XGBoost is a powerful boosting algorithm that can handle complex relationships and is faster than other boosting algorithms.

Results



1 XGBoost

R Square: 0.73

MSE: 0.24

2 Random Forest

R Square: 0.51

MSE: 0.35

3 Linear Regression

R Square: 0.27

MSE: 1.25

Post Modeling Analysis

The factors having the **greatest impact** on the rental prices based on XGBoost Modeling

1

Neighborhood

Different neighborhoods may have different characteristics, such as safety, accessibility, or amenities, that can affect the price.

2

Maximum Occupancy

Properties with higher maximum occupancy may be more popular among guests or have higher revenue potential.

3

Property Types

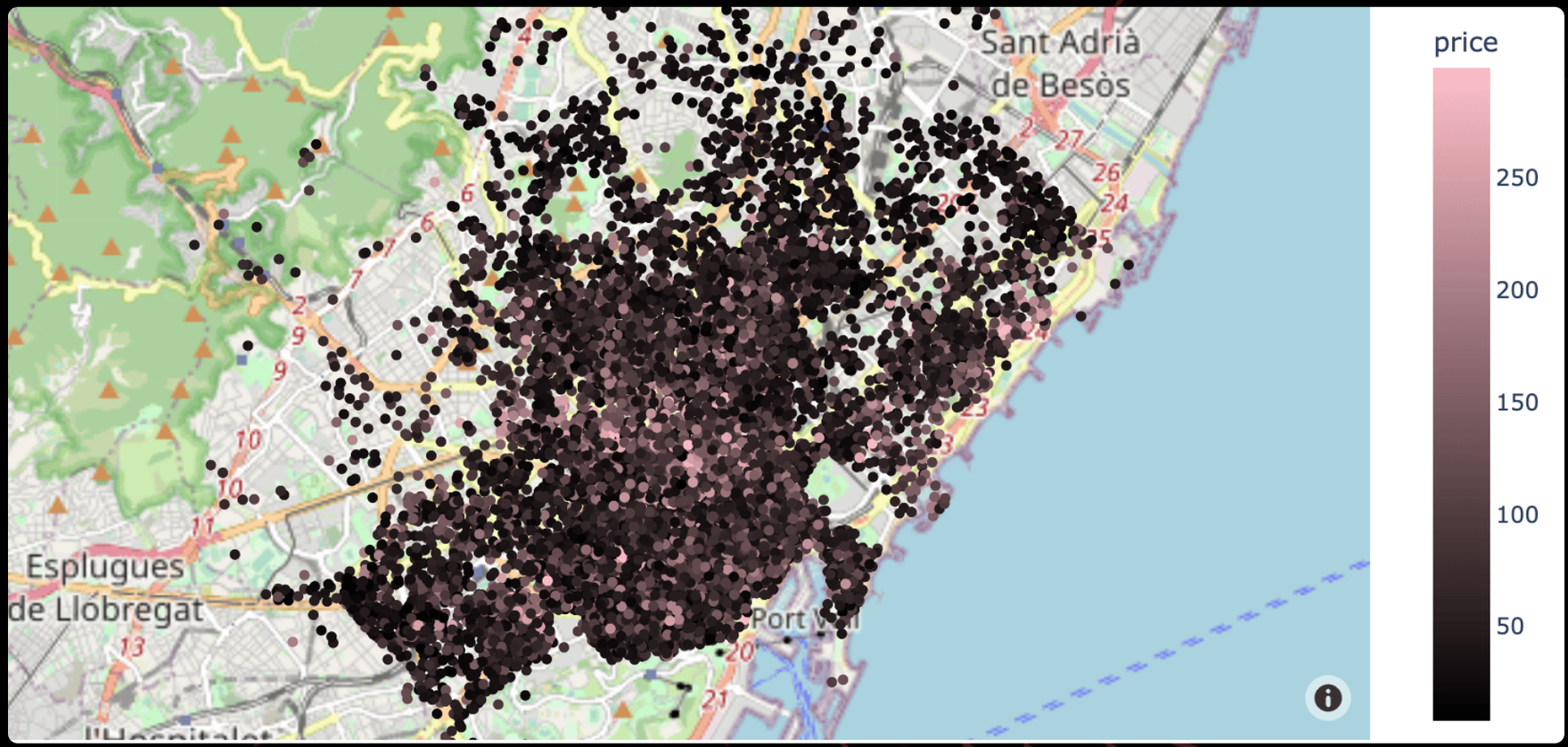
Different types of properties (e.g., apartments, lofts, condominium) have different characteristics that may influence the price.

4

Room Types

Different types of rooms (shared, private room, entire space) may have different amenities or pricing structures, which can affect the price.

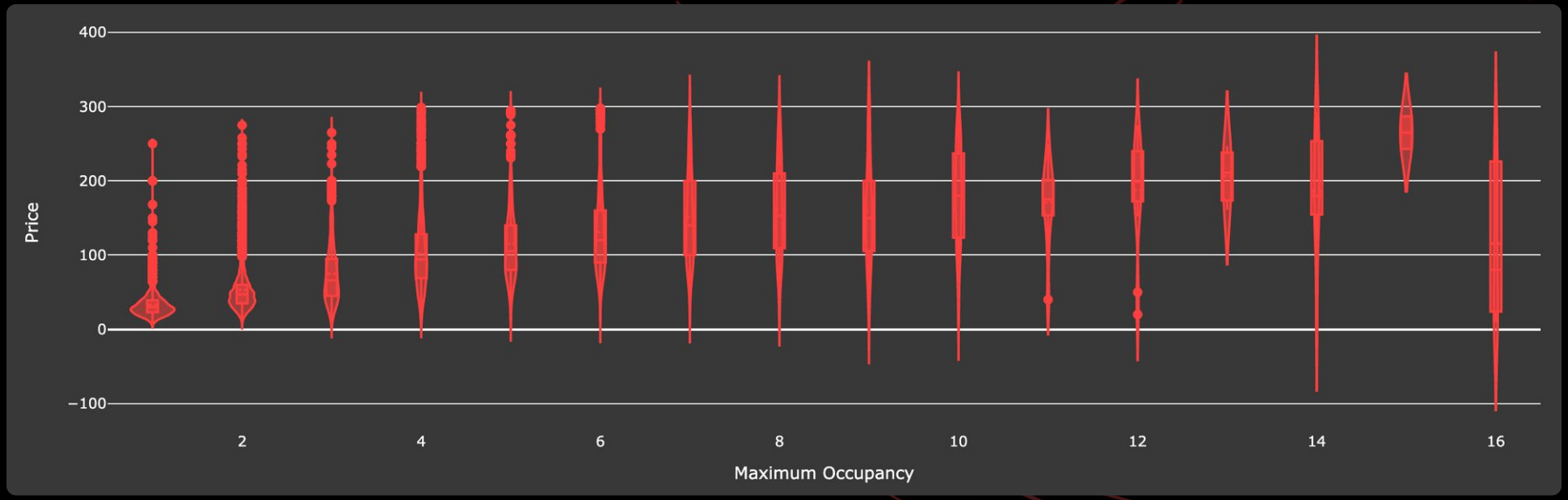
Neighborhood



The map reveals that rentals located in centrally located neighborhoods tend to have higher prices against neighborhoods in the outskirts of the city. This pattern can be explained by the fact that centrally located neighborhoods typically offer easier access to amenities, entertainment, and transportation, which makes them more desirable and thus commands a higher price.

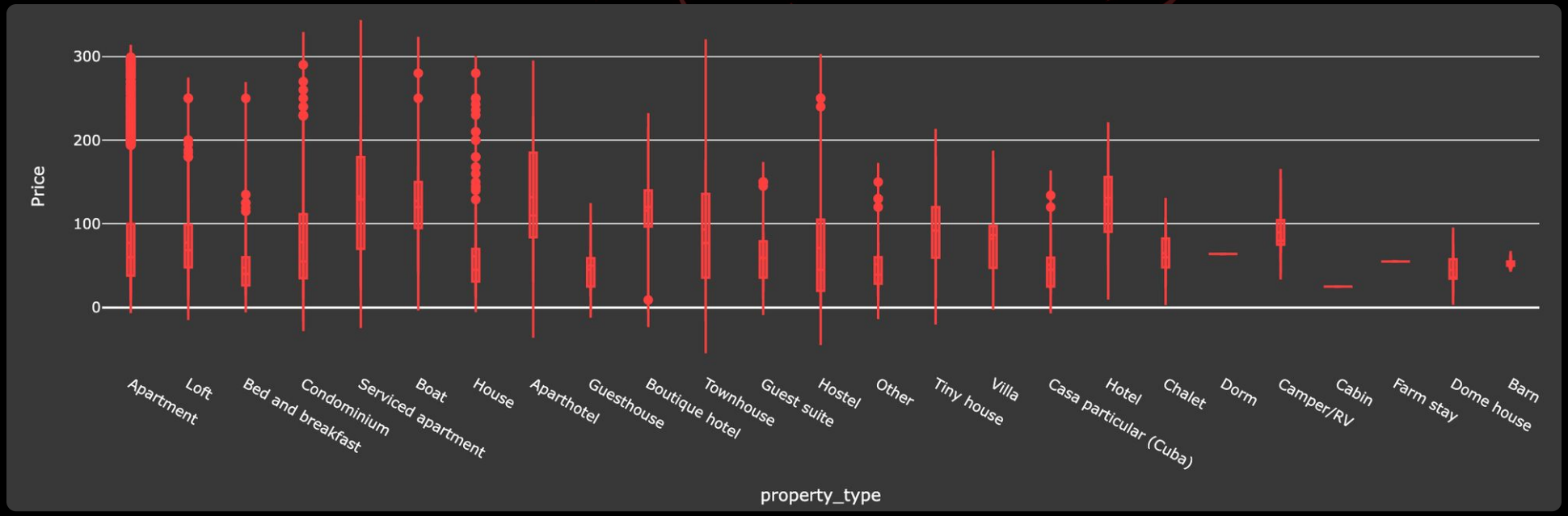
On the other hand, rentals on the outskirts of the city tend to have lower prices. This could be due to the fact that they are further away from the city center, which may make them less accessible and less desirable to some potential guests.

Maximum Occupancy



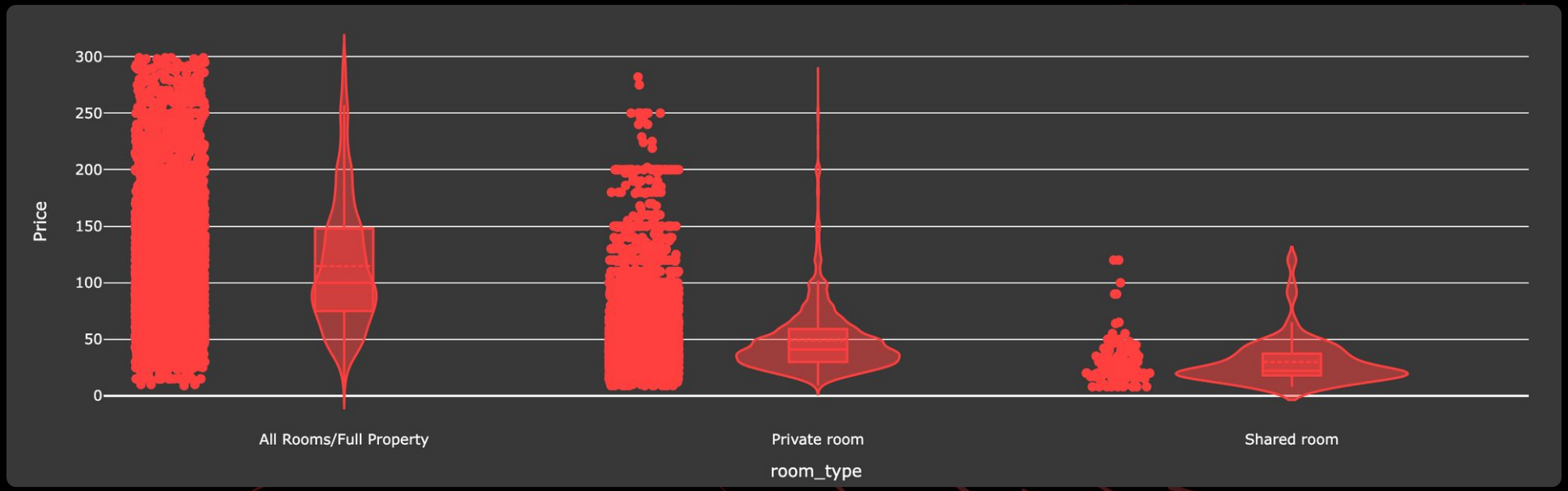
By analyzing this violin plot, property owners can better understand the relationship between maximum occupancy and price, and use this information to make data-driven decisions. For example, owners may adjust their pricing strategy based on the occupancy level they wish to target, while potential guests can use this information to compare prices across properties with different maximum occupancy levels.

Property Types



This violin plot can be explained by the fact that different property types offer different amenities, sizes, and locations, which can affect the cost of the property. For instance, lofts and condominiums may have more upscale finishes and amenities compared to guest houses or B&B, which can justify their higher prices.

Room Types



The violin plot reveals that full properties, which are properties that can be rented out entirely, tend to have the highest prices among the three room types. Private rooms, which are rented out individually but share common areas with other guests, generally have lower prices than full properties, but higher prices than shared rooms. Shared rooms, which involve sharing the same sleeping area with other guests, usually have the lowest prices among the three room types.

Summary

The above slides help us to highlight the various aspects of data exploration, modeling, and analysis of a Airbnb rental dataset. Exploratory data analysis is crucial for identifying important features and cutting down the number of irrelevant columns. Statistical modeling, such as linear regression, random forest, and XGBoost, can be used to predict property prices and determine the most important factors. Post-modeling analysis can provide insights into how different factors affect property prices, such as neighborhood, property type, room type, and maximum occupancy.

Conclusion

Ultimately, these insights can be used by new hosts looking to enter the market or existing hosts to make data-driven decisions based on their budget and preferences.

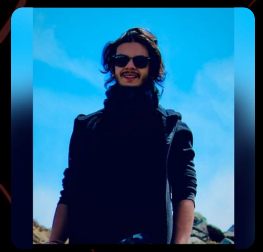
Meet the Team...

Devansh Singh

DPS220001

Dimple NC

DXN210024



Sarthak Vajpayee

SXV220020

Yash Agrawal

YXA220007