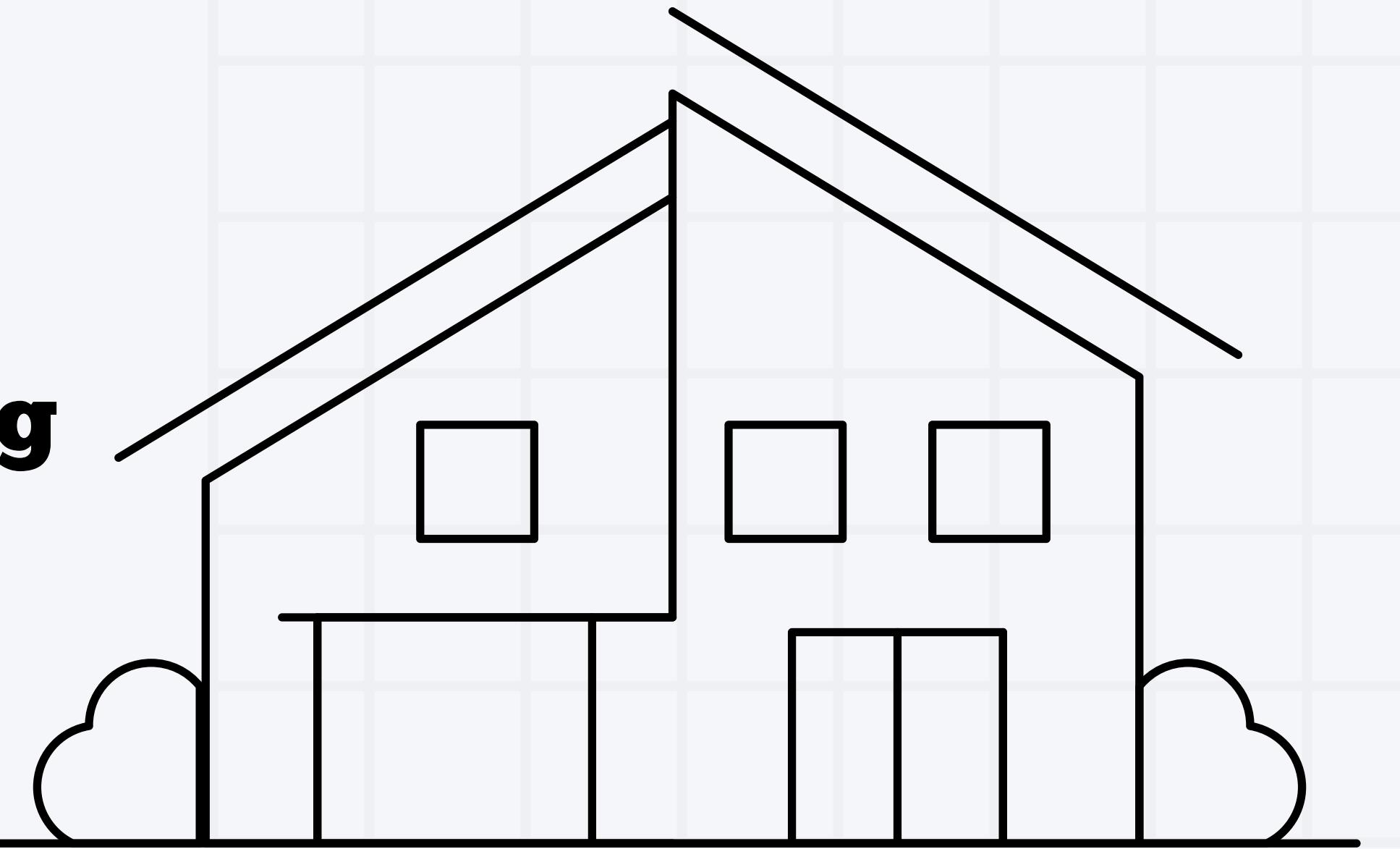


# **Melbourne Housing Market Prediction using CatBoost**

Presented by Shaltsa Nadya





# PROJECT TIMELINE

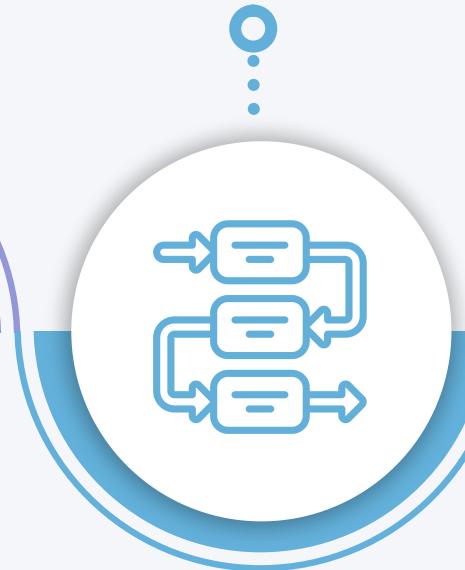
## Data & Methodology

- Raw data provided
- Variables in the dataset
- Flow diagram



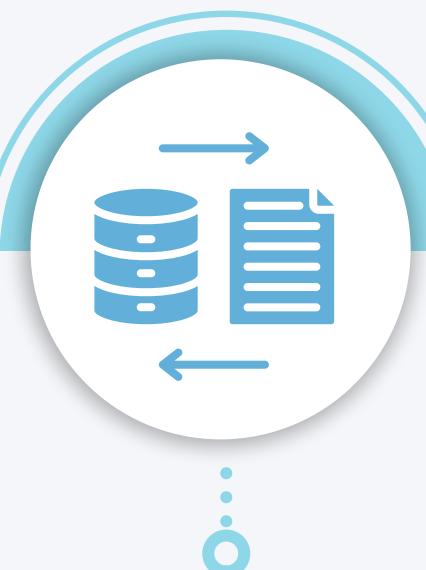
### Introduction

- Melbourne housing market context
- Project Overview
- Key-objective of the project



### Data Pipeline

- Data Ingestion & Data Transforming
- Data Analysis



## Data Analysis

- Data preprocessing
- Propose Dashboard



### Model Development

- Machine learning model
- Model evaluation



## Result & Discussion

- Business Integrity
- Summary of the findings



# Introduction

## Competitive Environment ?



The more modest movements in Melbourne dwelling values likely comes down to more supply of dwellings over the past 15 years.

- Less competition for more **affordable properties**
- Many people can't afford **freestanding homes**
- Houses built in the 1970s** are popular with small families

### Median price snapshot for houses

Median price ⓘ

**\$813,500**

December 2022 - November 2023

Past 12 month growth ⓘ

**Up 14.8% ↑**

### 5 year median price trend



The economic fallout of the COVID-19 pandemic, extended lockdowns, consecutive interest rate hikes, and tighter lending restrictions to name a few – and a trough-to-peak growth of 17.3% **during the 2020-21 property boom**, **Melbourne's housing values appear to have turned a corner**.

## Company Strategy



What type of house sells best?



Analyze **price trends** based on house type, location and facilities.

## Predict Price

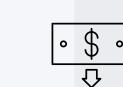


Analyze the needs

## The Challenges



Demand and Supply



Economic Conditions (Inflation)



Demographic and Lifestyle Changes

**Source**

# Framework Tools



**Powerful** analytics for everyone. Analytics innovations help anyone—from data scientists to business users—uncover insights faster.

→ **Flexibility and choice**  
→ **Integration and extensibility**

# Streamlit

An open-source Python framework for **machine learning** and **data science** teams

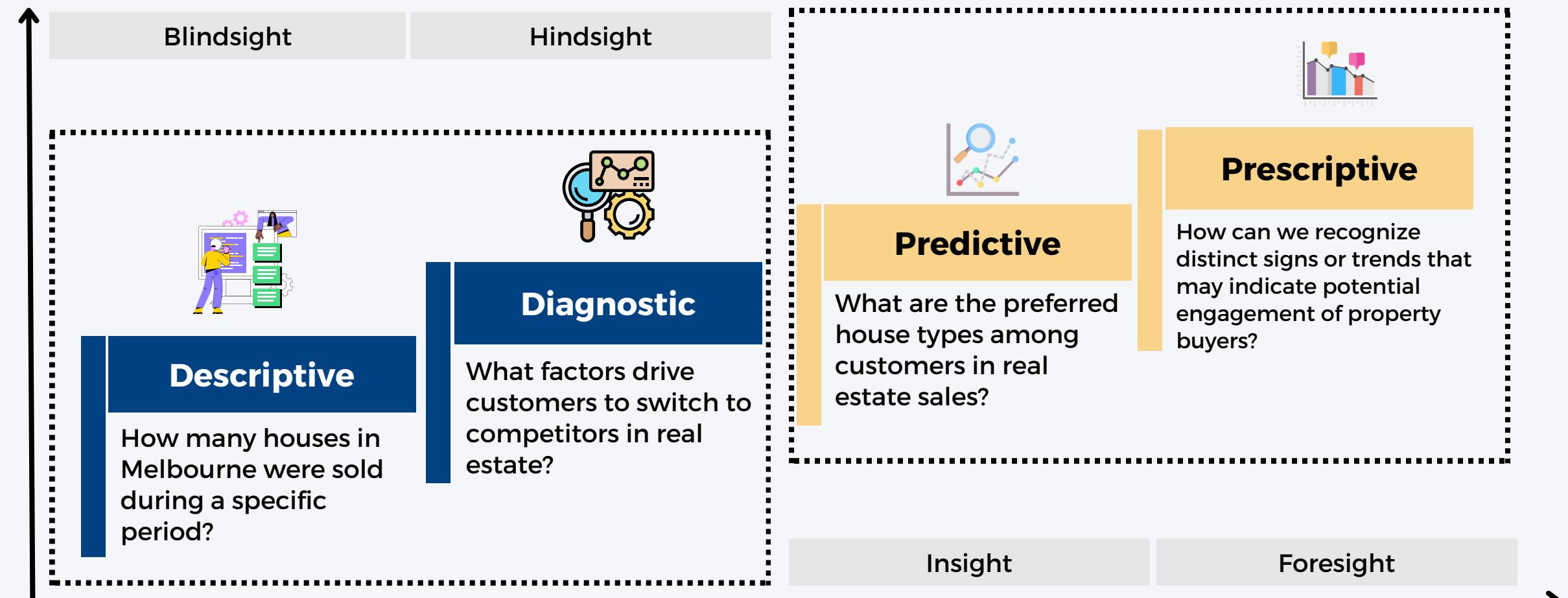
→ **Deploy instantly**

# PostgreSQL

an Object-Relational Database Management System (ORDBMS) with **Robust feature set**

→ **Support SQL, Python , R, etc**

- Define your own complex data types
- Overload functions to work with different argument data types
- Define inheritance relationships between tables



## Data Analysis



Define Business



Gather Data



Dashboard



Formulate  
Business Strategy



Strategy  
Implementation



Business Strategy  
Recommendations

## Machine Learning



Extract Data



Clean and  
prepare data



Identify needs



Create Model



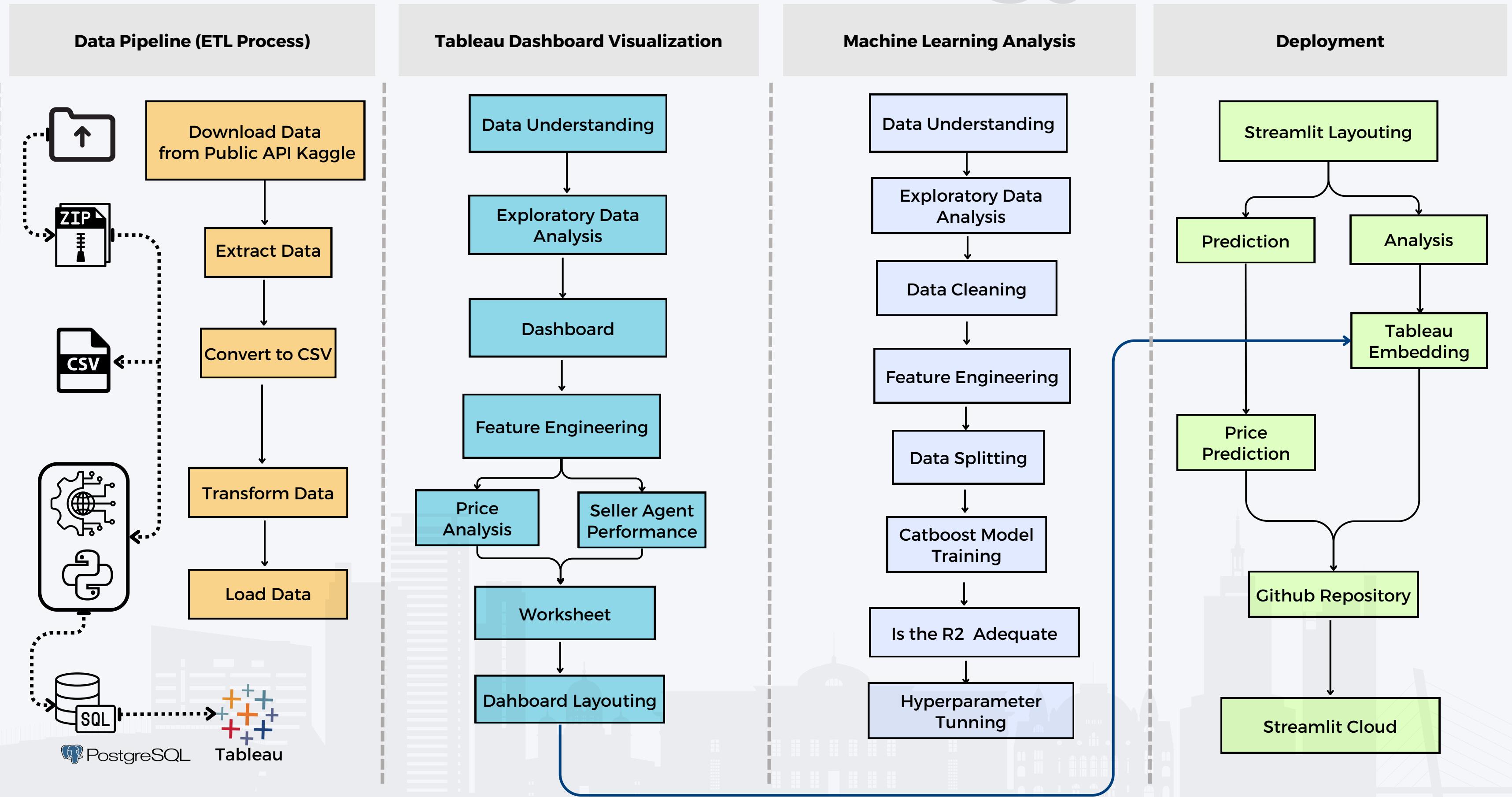
Prediction



Develop plan

→ **Output**

# Methodology



## Top Performance Seller Agent for each Type of House

```
top_perform as (
    select type, (select sellerg
        from (select sellerg, count(seller)
            from melbourne
            group by 1
            order by 2 desc)
        limit 1),
        count(seller) as qty_sold,
        count(seller)*100/(select count(*)
            from melbourne)
        as percentage
    from melbourne
    group by 1
    order by 3 desc),
```

### Result

	type character varying (5)	seller character varying (55)	qty_sold bigint	percentage bigint
1	h	Nelson	9448	69
2	u	Nelson	3017	22
3	t	Nelson	1114	8

### Sales Rate %

```
select
round((sum(qty_sold17)-sum(qty_sold16))
    *100/sum(qty_sold17),2) as percentage
from growth
```

	percentage numeric
1	11.13

**Insight:**  
Demand for Melbourne housing stock is likely to **11.13%**

## Growth rate from 2016-2017

```
sold16 as (
    select
        seller,
        count(seller) as qty_sold16
    from melbourne
    where date_trunc('year', date)= '2016-01-01'
    group by 1
    order by 2 desc),
    sold17 as (
        select
        seller,
        count(seller) as qty_sold17
    from melbourne
    where date_trunc('year', date)= '2017-01-01'
    group by 1
    order by 2 desc),
```

```
growth as(select sold17.seller,qty_sold16,qty_sold17,
Round((qty_sold17-qty_sold16)*100/qty_sold16,1) as growth_rate
from
sold17
inner join
sold16
on
sold17.seller=sold16.seller
order by 3 desc
)

select *
from growth
```

## The Most Improved Seller Agent

	seller character varying (55)	qty_sold16 bigint	qty_sold17 bigint	growth_rate numeric	most_improved bigint
1	Ray	229	472	106.0	243
2	Barry	423	588	39.0	165
3	YPA	18	136	655.0	118
4	Harcourts	29	139	379.0	110
5	McGrath	72	150	108.0	78

## The Worst Sales of Seller Agent

	seller character varying (55)	qty_sold16 bigint	qty_sold17 bigint	growth_rate numeric	most_improved bigint
1	Nelson	867	698	-19.0	-169
2	RT	113	71	-37.0	-42
3	Miles	115	81	-29.0	-34
4	Marshall	346	313	-9.0	-33
5	Brad	186	156	-16.0	-30

### Insight:

Seller agents who were stuck or did not experience an increase in sales and even had their performance decline from 2016 to 2017

# Exploratory Data Analysis (EDA)

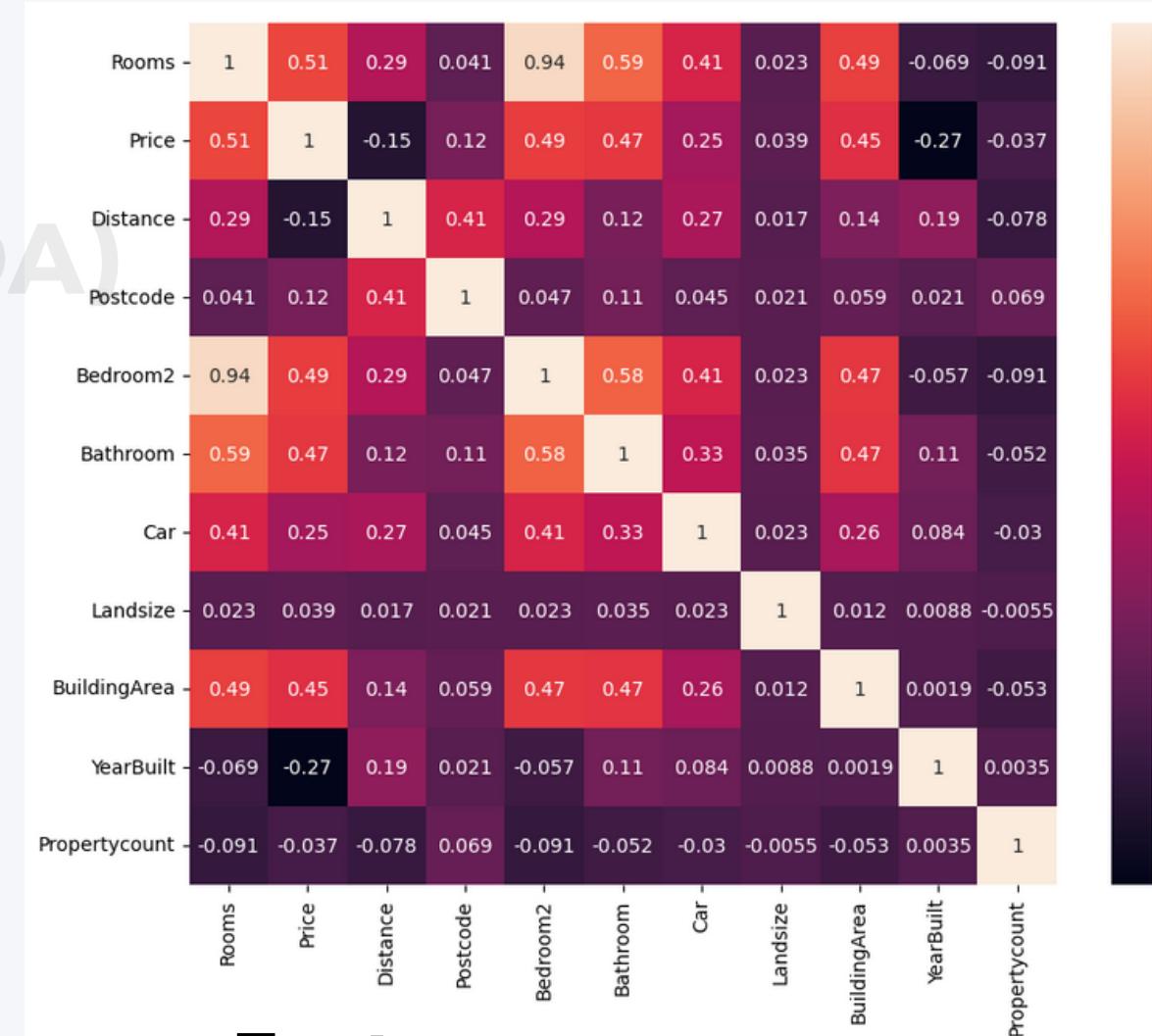
The heatmap graph indicates a strong correlation about to **0.94** between the variables **bedrooms and rooms**, suggesting that as the number of bedrooms increases, there is a higher likelihood of having more rooms in the property.

## Melbourne House Dataset Preview

**22**  
Columns

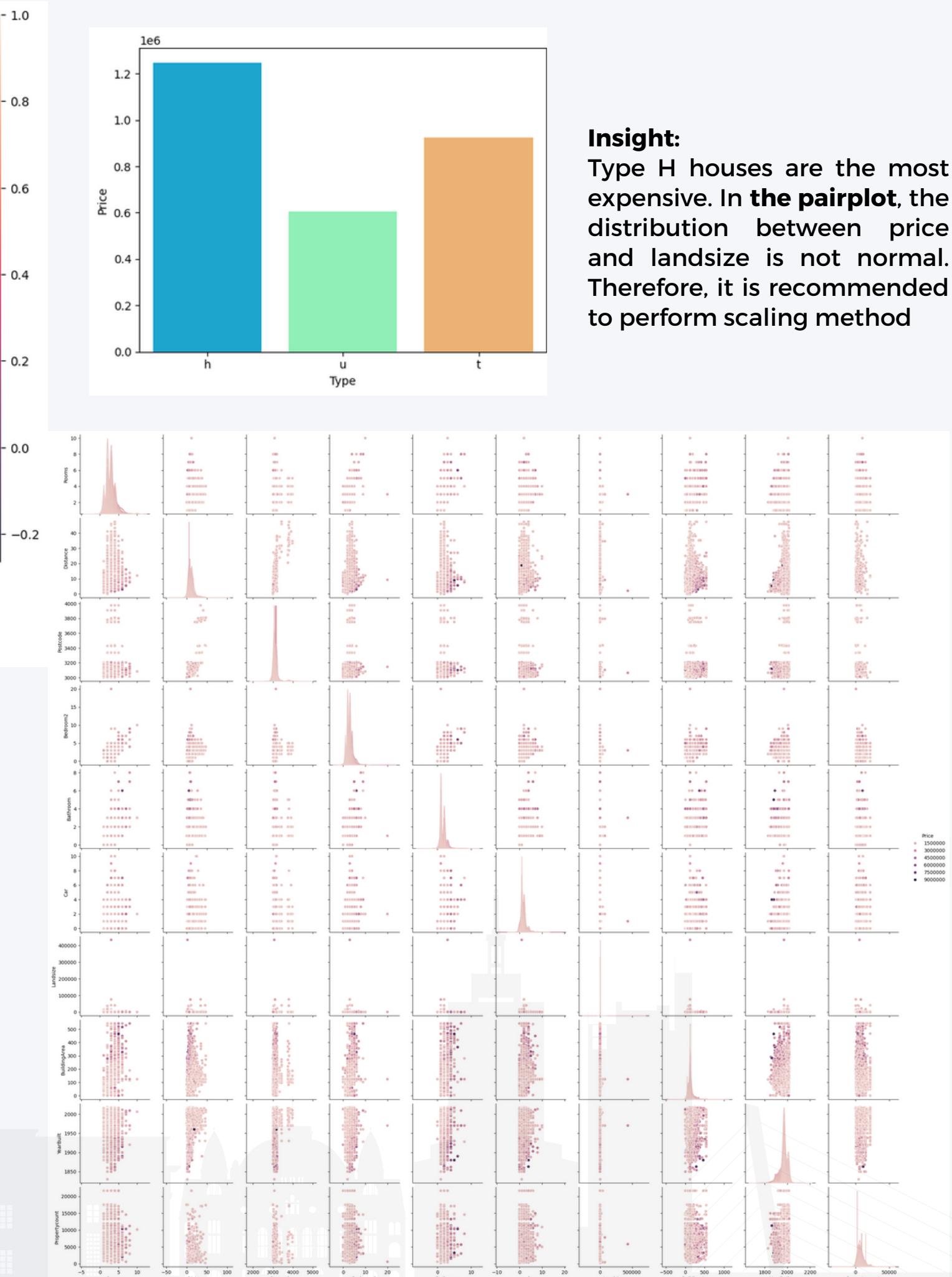
**13,580**  
Total Houses Sold

**85,000-9,000,000**  
Range Price (AUD)



## Features

- 0 Suburb
- 1 Address
- 2 Rooms
- 3 Type
- 4 Price
- 5 Method
- 6 SellerG
- 7 Date
- 8 Distance
- 9 Postcode
- 10 Bedroom2
- 11 Bathroom
- 12 Car
- 13 Landsize
- 14 BuildingArea
- 15 YearBuilt
- 16 CouncilArea
- 17 Latitude
- 18 Longitude
- 19 Regionname
- 20 Propertycount



### Insight:

Type H houses are the most expensive. In the pairplot, the distribution between price and landsize is not normal. Therefore, it is recommended to perform scaling method

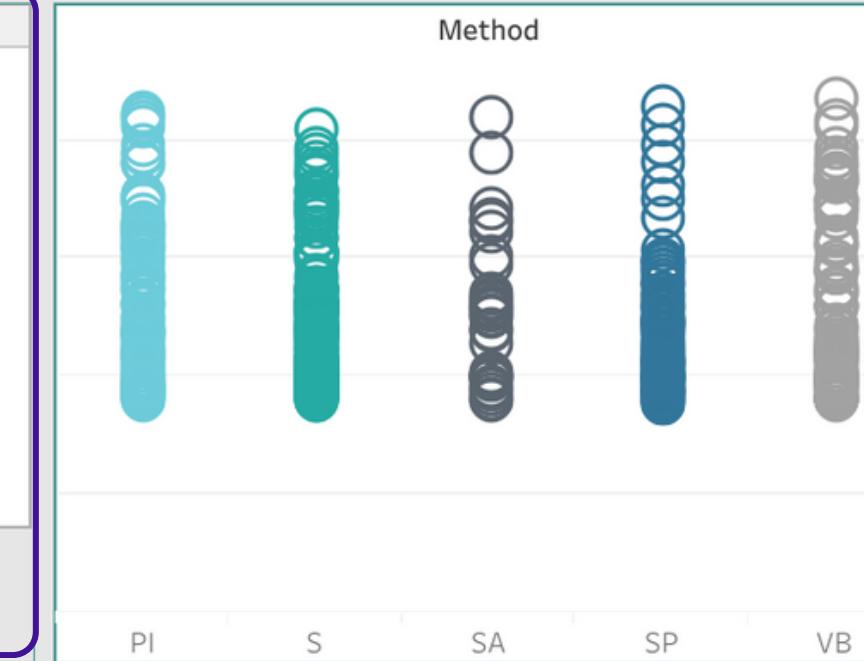
# Melbourne House Market Dashboard

# Data Analysis

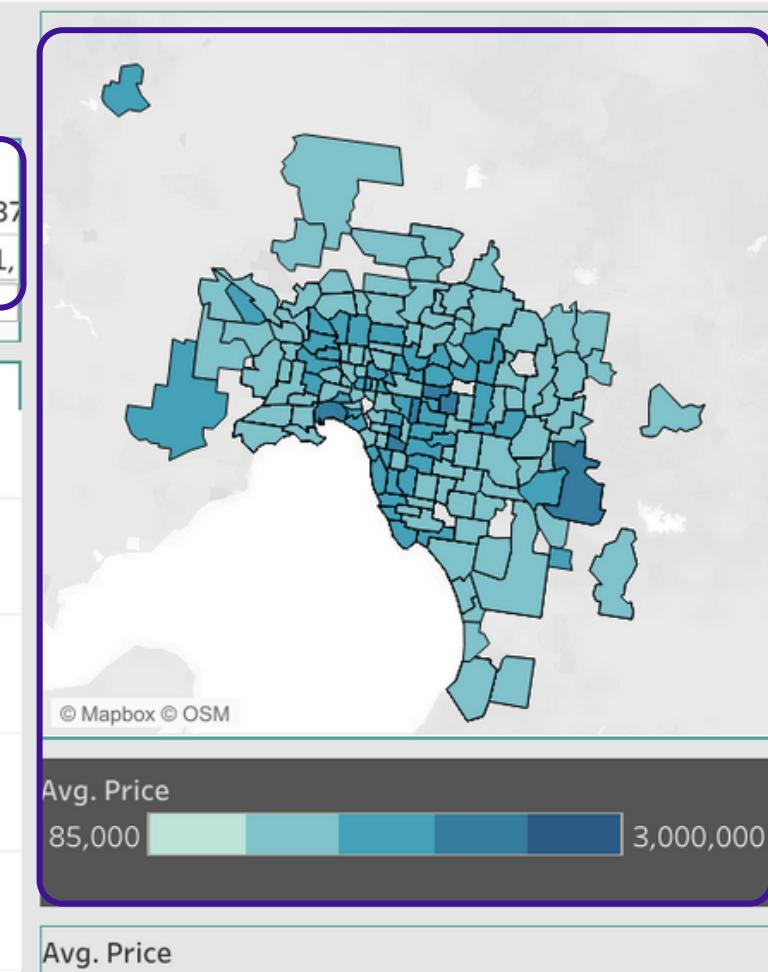
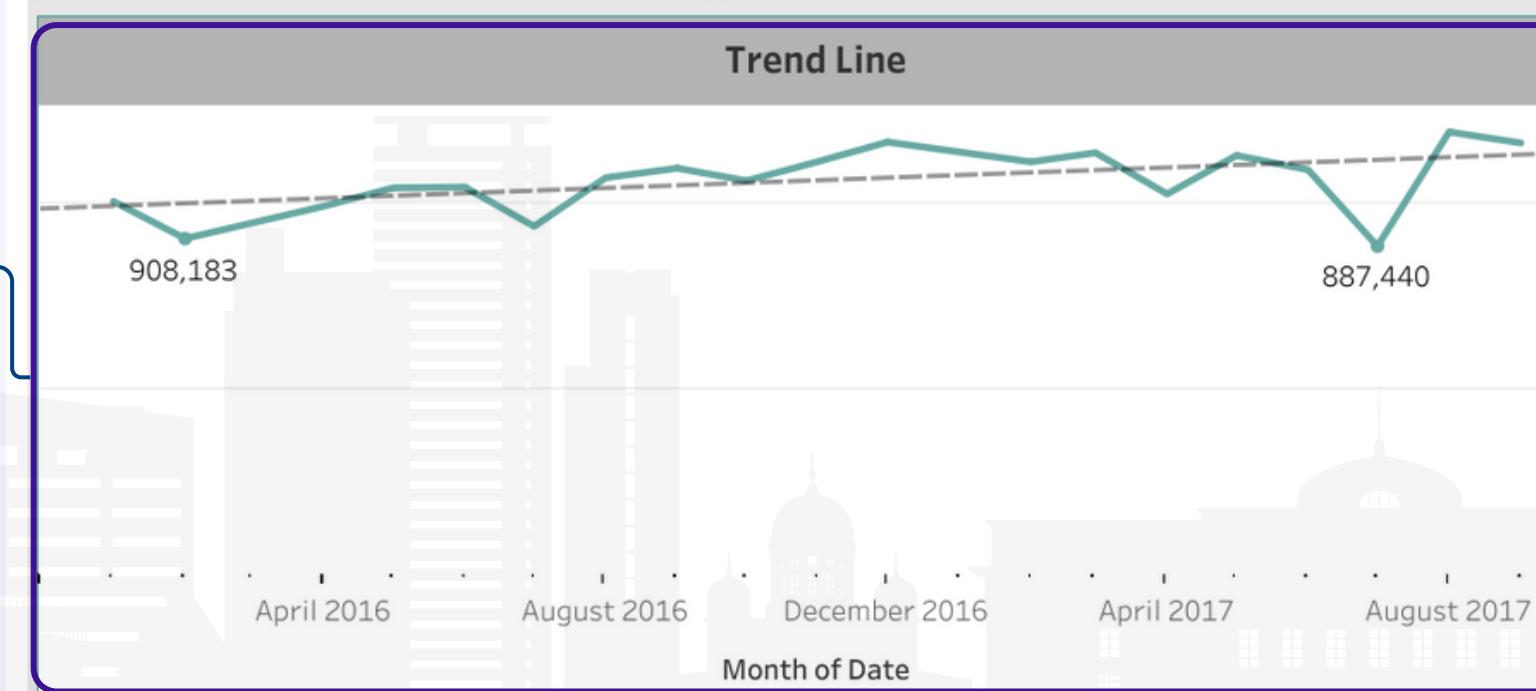
1 Average house price in each year it was built



2 Group by average house price in each suburban area



3 House price trends in Melbourne based on the time series show a considerable increase.



Avg. Price  
887,440.225260417 to 2,185,000

Type  
 h  
 t  
 u

Method  
 PI  
 S  
 SA  
 SP  
 VB

5 Geo Map based on postcode, suburb component with color gradations of house prices

4 select button by house type and payment method

# Model

## Model Development



a powerful open-source Python library designed for data manipulation and analysis. Supports various data formats, including CSV, Excel.



provides a comprehensive set of tools for tasks such as classification, regression, clustering, and dimensionality reduction. include model evaluation tools.



CatBoost

a high-performance machine learning library designed for gradient boosting on decision trees. Handling categorical features without the need for extensive preprocessing.



a versatile Python library widely used for data visualization. It provides a flexible and comprehensive set of tools

Metrics

R2 Score

**0.78**

MAE

**187.852,25**

- Imported the dataset
- Dropped unnecessary columns such as 'Latitude', and 'Longitude'.
- Renamed certain columns for easier understanding.

Data Cleaning

Feature Engineering

Model Training

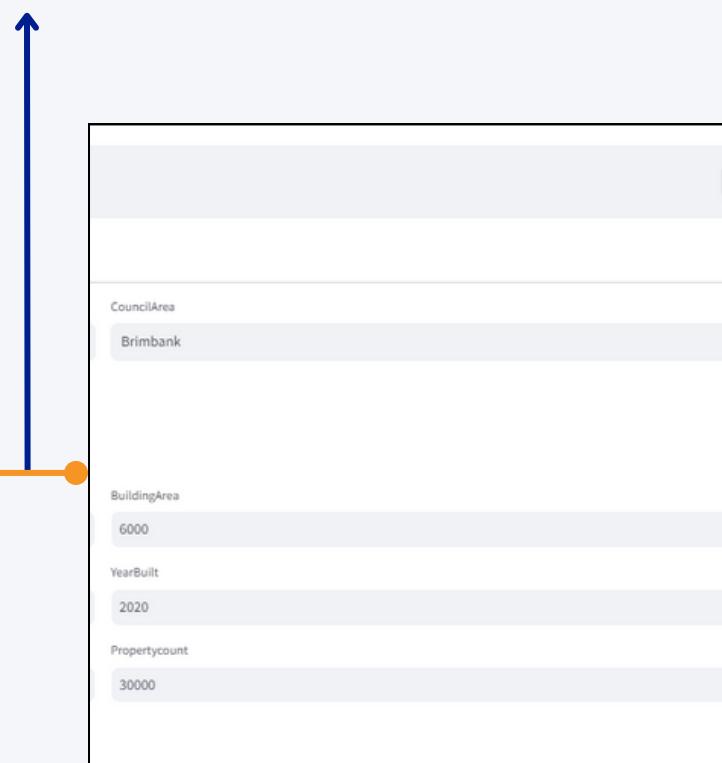
Model Evaluation

- Scaling on Numerical Variables using **scikit-learn's**
- Transform Categorical Variables with **Encoder Labels** like 'Method', 'Type', 'CouncilAreal', and 'Regionname'

- **CatBoostRegressor** model to predict 'Price' using (iteration, learning rate, depth parameter)
- **Split** the data using scikit-learn's
- **Train\_test\_split** function.
- Fit the model on the Training data

- Made **predictions** on the test set.
- Evaluated the model's performance using the mean absolute error (**MAE**) and **R2 Score** metric

## Variables related to the condition of the houses for sale



1 Customize the company needs based on newest datasets to estimate and predict price house based on building characteristic

Analyzing the housing market conditions in Melbourne to facilitate the development of a business strategy based on geographical factors and building characteristics

2

## PREMIS

### Prediction of Real Estate Market in Dashboard Analysis

Real Estate Analysis

Prediction

Model Evaluation

#### Model Evaluation

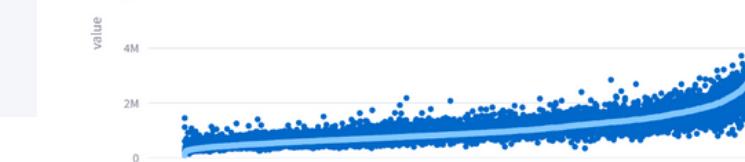
MAE  
187852.25

MSE  
89203636198.29

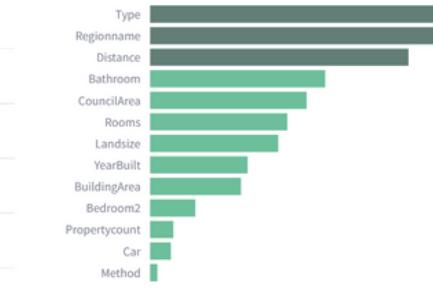
R2 Score  
0.78

#### Predicted House Price vs House Price

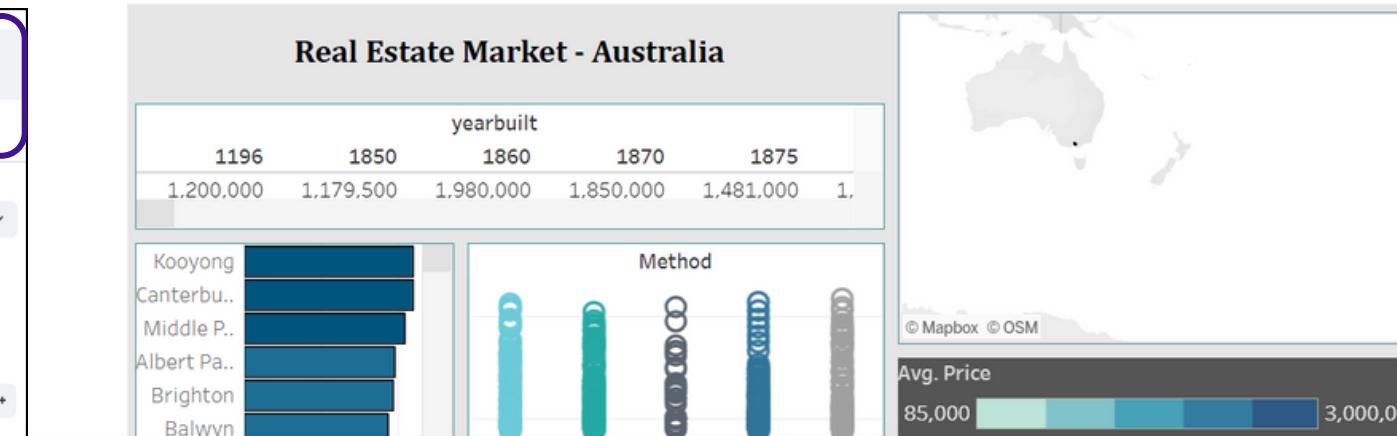
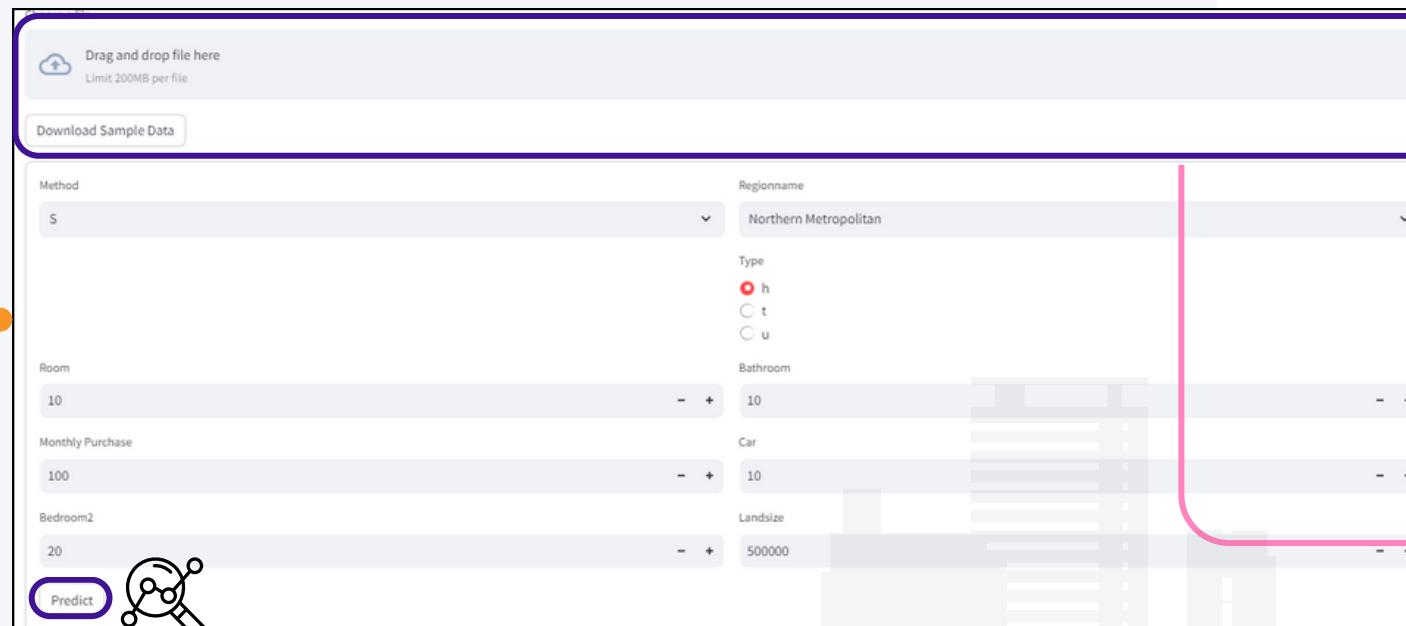
variable  
Predicted Price  
House Price



#### Feature Importance



3 Making appropriate decisions for the data team is a critical component of dealing with severe risks.



4 Enable uploading data in the form of CSV or Excel files and load the data for predicting house prices. Additionally, include a download button that allows users to download sample data in the expected format.

<https://premis.streamlit.app>



What is the **growth rate** from 2016 to 2017?

Which **areas do customers prefer** to buy houses, with an average price?

what **type** of houses are **most favored**?

What is the **trend in prices** based on historical data?

# Conclusions

## Conclusions

1. The growth rate from 2016 to 2017 is **11.13%**.
2. In terms of **preferred locations**, the suburb of **Kooyong** stands out, with an average house price of 2,185,000 AUD. On the other hand, **Rowville** has the **lowest average** price at almost 890,000 AUD.
3. The **price trend** shows a **gradual increase**, although not significantly, from month to month.



**For further inquiries or collaboration discussions, contact me directly:**



Email: [shaltsa.nadya99@gmail.com](mailto:shaltsa.nadya99@gmail.com)



<https://premis.streamlit.app>



<https://github.com/shaltsanadya>



<https://linkedin.com/in/shaltsanadya>