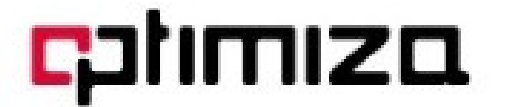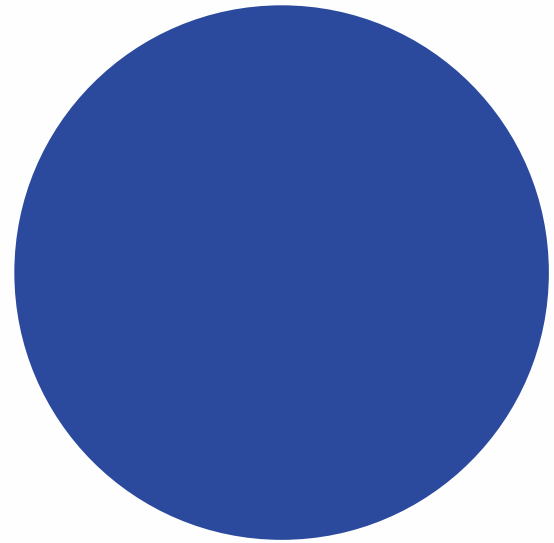Data Science Graduation Project

KING COUNTY

Data driven insights and predictive modelling tool

Hamza Jahamah

Feras Al-Qirnawi

Abdullah Khader
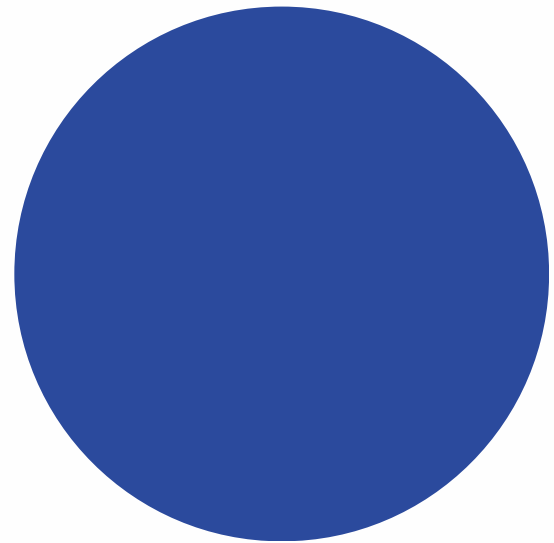
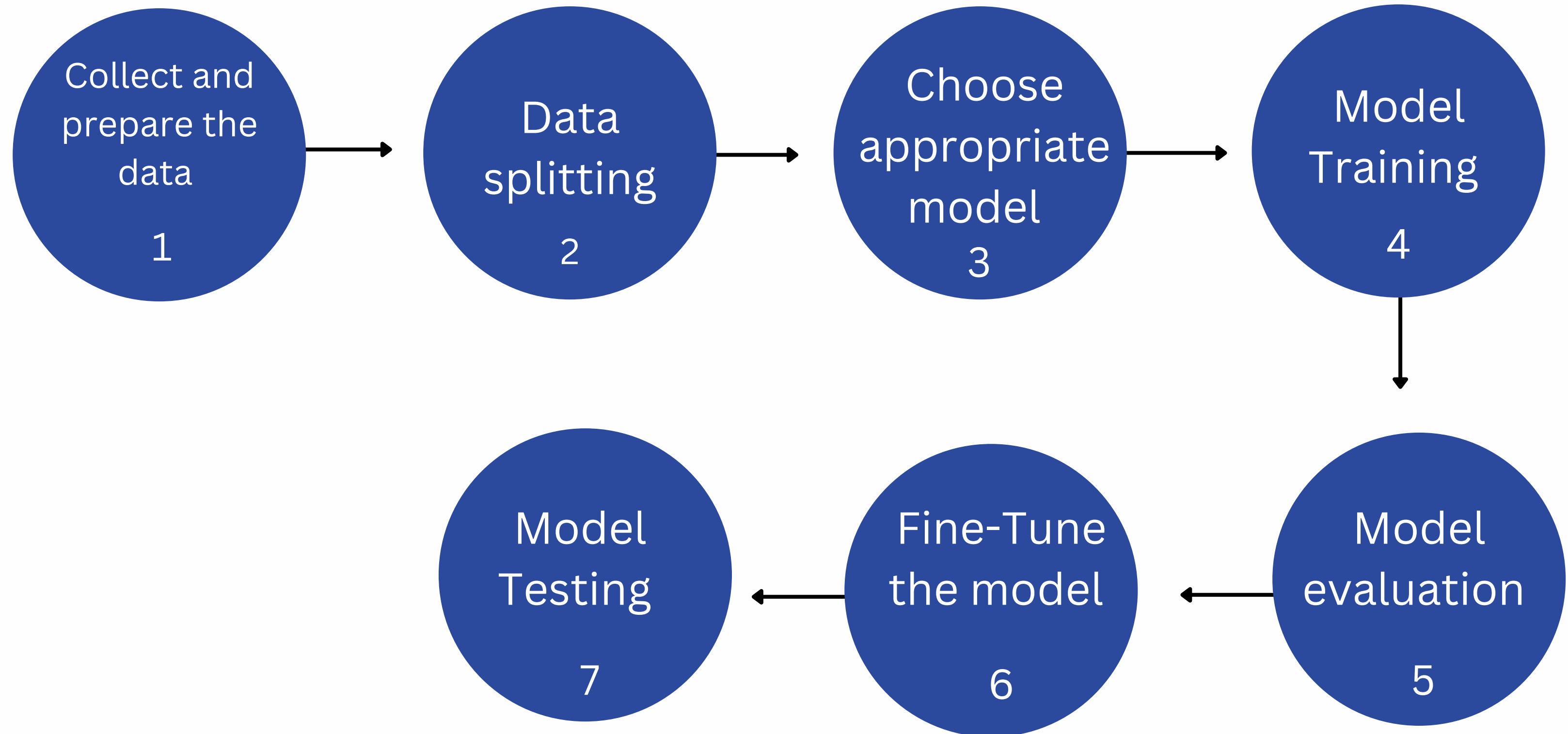# PROBLEM STATEMENT

HOUSE SALES IN KINGS COUNTY

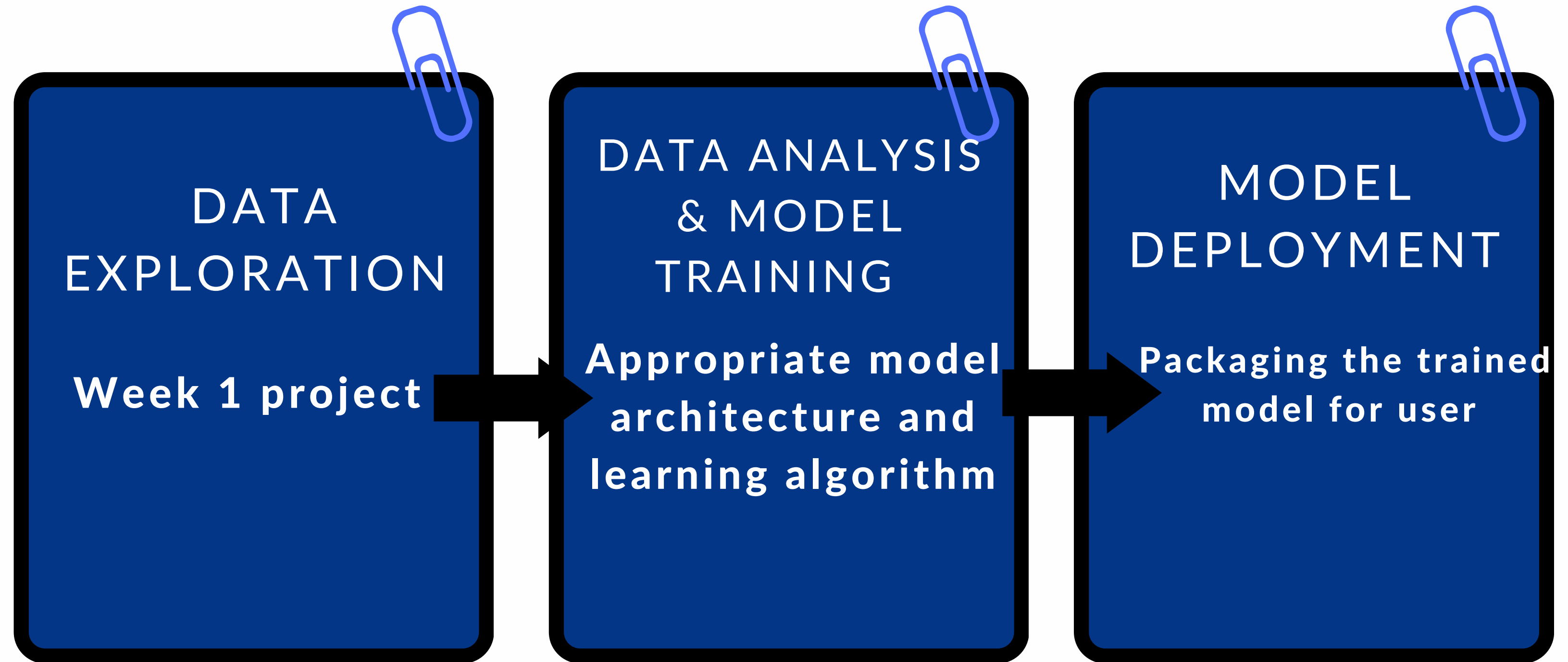- **Data driven recommendations**
- **Model to predict house prices**

GOAL

- **Train and validate a prediction model to accurately predict the sale price of houses in kings county .**

APPROACH

DATA EXPLORATION

Week 1 project

DATA ANALYSIS & MODEL TRAINING

Appropriate model architecture and learning algorithm

MODEL DEPLOYMENT

Packaging the trained model for user

SELECTED HOUSE FEATURES

1- No of bedrooms
2- No of bathrooms
3- Square-feet living area
4- Square-feet lot area
5- No of floors
6- House waterfront
7- No of times viewed

8- House condition
9- House grade
10- Square feet basement area
11- Year built
12- Year renovated
13- Lateral coordinate
14- Longitude coordinate

PREDICTING HOUSE PRICES

	Model 1 Linear Regression	Model 2 Decision Tree	Model 3 Random forest	Model 4 CatBoost
Features	14	14	14	14
Score Test score	66%	80%	83%	88%
Mean Error	115,000 USD	88,000 USD	71,000 USD	49,000 USD

PREDICITION MODEL TOOL INTERFACE

FUTURE IMPROVMENTS

1 INCREASE THE AMOUNT OF DATA USED TO TRAIN THE MODEL

The more data the model has to learn from , the more accurate it will be.

2 INCORPORATE EXTERNAL DATA

In addition to data on past home sales, it may be useful to incorporate other types of data that could affect house prices, such as economic indicators or natural disasters.

3 MONITOR AND UPDATE THE MODEL REGULARLY

As the housing market changes over time, it will be important to periodically re-train the model to ensure it remains accurate.

Thank you 😁

Any Questions ??