User cases

Exploratory data analysis

Buyers	1 Finding potential suburbs for investment with a fixed range of budget 2 Finding the best timing to enter housing market/which month has the best sale performance 3 Comparing the changes of housing median price for each month suburbs 4 Identifying popular suburbs for investment 5 Identifying the price change along with distance to CBD 6 Comparing 16 vs 17 housing median price in different suburbs
Sellers	1 Identifying the best sale month/season, finding the best timing to sell a house 2 Finding good agents (who can sell top prices and sell more houses than other agent) 3 Finding the best/suitable method to sell a property (unit or house)/popular way of selling a house 4 Exploring neighbouring market, identifying the house price within my suburb to offer a good price
Real Estate Agent	1 Having better understanding of their competitors, knowing their strategies, how many houses do they sell, in which suburbs 2 Identifying the sale hot spot, which suburbs have the best sale performance
Developer	1 Identifying the best sell house type, the most suitable land size, house size, rooms etc. for development 2 Identify the hot spot for investment
Government	1 Identifying the suitable house type, the most suitable land size, house size, rooms etc. for development 2 Identifying the potential needs for housing in different suburbs

Predictive data analysis

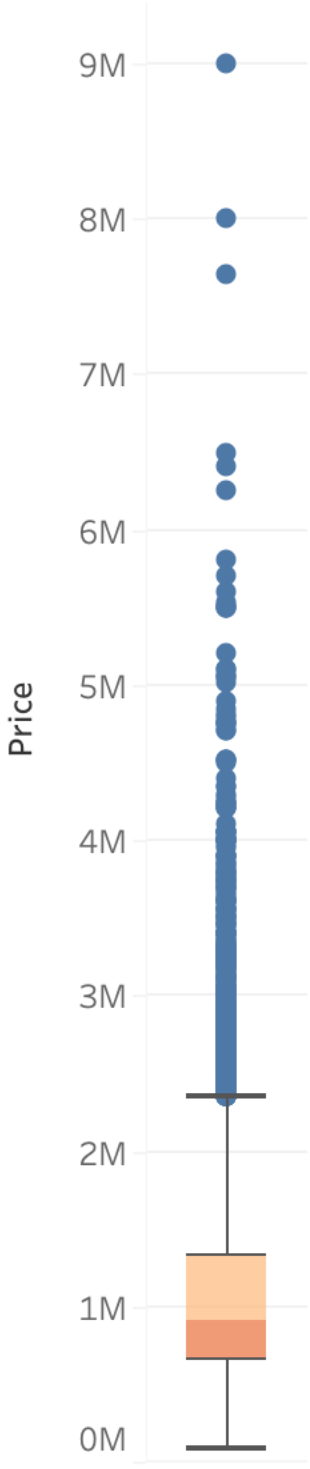
1 Predict the price of a property
2 Predict the sales performance in the next coming season
3 Predict the best growth area

Data cleaning

1 Missing data

Car	62	Fill missing data with Mode of Car
BuildingArea	6450	Fill missing data with Mean of BuildingArea
YearBuilt	5375	Fill missing data with Mode of YearBuilt
CouncilArea	1369	Ignore these values since it is not considered in this use cases

2 Noisy data



Outliers have been detected from the Box plot, those values above the upper quartiles Need to be removed.

	Price	Rooms	Distance	Landsize	BuildingArea	PropertyCount
Mean	1075684.079	2.937997054	10.13777614	558.4161267	151.9676499	7454.417378
Standard Error	5486.0796	0.008201504	0.050360945	34.24489577	6.407139359	37.5736668
Median	903000	3	9.2	440	126	6555
Mode	1100000	3	11.2	0	120	21650
Standard Deviation	639310.7243	0.955747938	5.868724943	3990.669241	541.0145376	4378.581772
Sample Variance	4.08718E+11	0.913454122	34.44193246	15925440.99	292696.7299	19171978.33
Kurtosis	9.874338886	0.794067989	5.260001109	10180.34683	6347.802222	1.217820011
Skewness	2.239624313	0.376478033	1.676937083	95.23740045	77.69154092	1.069339349
Range	8915000	9	48.1	433014	44515	21401
Minimum	85000	1	0	0	0	249
Maximum	9000000	10	48.1	433014	44515	21650
Sum	14607789799	39898	137671	7583291	1083529.344	101230988
Count	13580	13580	13580	13580	7130	13580

As shown in the Table above, the dataset covers 13580 transaction which is two-year data of Melbourne Properties from 2016 to 2017. Since the skewness values of all the columns are greater than 0, the distribution of data of all the columns are right skewed. Standard Deviation shows the statistical dispersion of the data, thus, most of the values of Price, Landsize and PropertyCount have great differences from their mean values while Rooms, Distance, BuildingArea do not have.

Feature Selection

Plot Parameter (Required Column of Data)					
Target users	**No.**	**X-axis**	**Y-axis**	**Plot type**	**New column created**
Buyers	1	Budget/price range	Suburbs	table/bar chart	Budget/price range
	2	Date(Month)	Count of Sales	Dual Lines	Date(Month), Count of Sales
	3	Suburbs	Median price	Bar chart	Median price of suburbs
	4	Suburbs/regions	Count of Sales	Bar chart/tree maps	
	5	Longtitude, Lattitude	Price	Maps	
	6	Date(Month)	Price	Dual Lines/bar chart	
Sellers	1	Date(Month)	Count of Sales	Dual Lines	
	2	SellerG	Count of Sales/price	Pie and bar chart	
	3	Method	Count of Sales/price	Pie and bar chart	
	4	Longtitude, Lattitude	Price	Maps	
Real Estate Agent	1	SellerG	Count of Sales/price/suburbs	Pie and bar chart	
	2	Suburbs/regions	Count of Sales	Bar chart/tree maps	
Developer	1	Type/LandSize/HouseSize	Count of Sales/price	Pie chart/bar chart	
	2	Suburbs/regions	Count of Sales	Bar chart/tree maps	
Government	1	Type/LandSize/HouseSize	Count of Sales/price	Bar chart/tree maps	
	2	Suburbs/regions	Count of Sales	Bar chart/tree maps	

New column created: Budget, Date(Month), Count of Sales, Median price of suburbs