

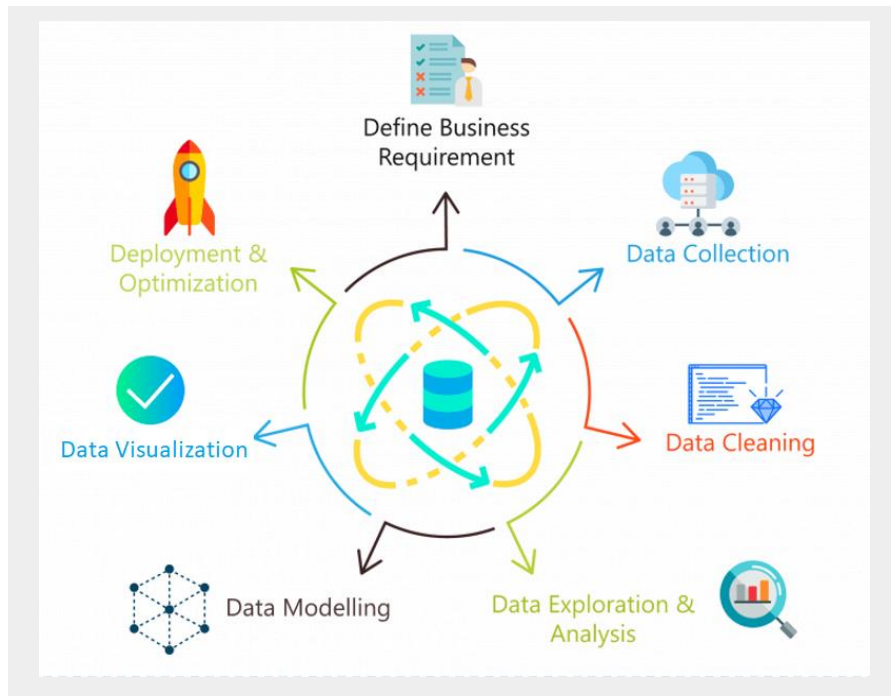
# **A brief summary of the first month & plans for next steps**

**-- Tricia**

# Agenda

1. A brief summary
2. Examples: SEEK and Melb\_data datasets
3. Planning: Melbourne Housing data analysis

# Data Analysis Process and What have been covered



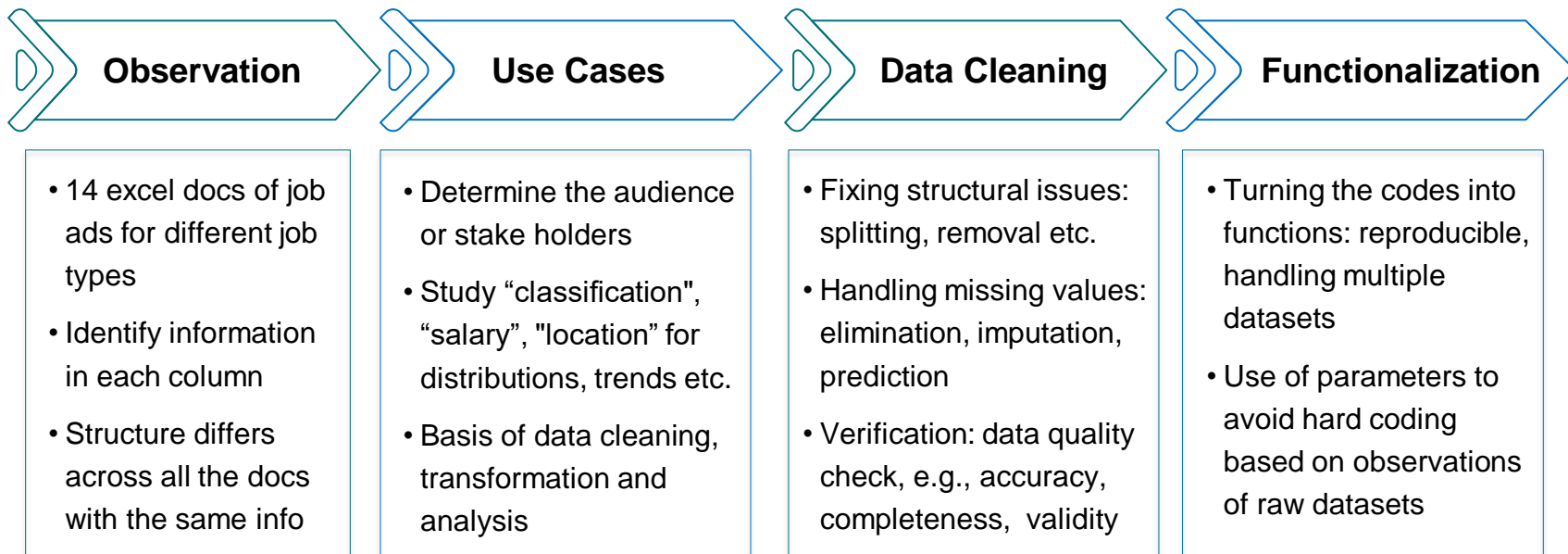
## 1. Get data ready: ✓

- **Use cases:** requirements/expectations of stake holders; a process of constant repetition and reinforcement
- **Data collection:** use datasets from SEEK and Melbourne Housing for practice
- **Data pre-processing:** ETL, data cleaning (missing value, outliers, data quality check, etc.)

## 2. Data exploration: ✓

- **Data exploration:** create a broad picture of important trends and major points to study in greater detail
- **Data visualization:** present a more straightforward view of datasets

# Example: data pre-processing for SEEK datasets

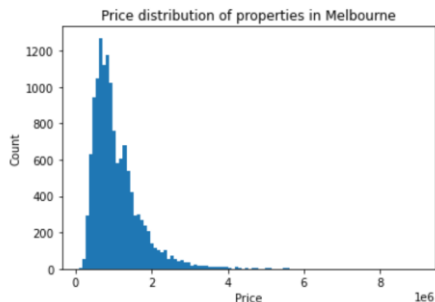


- ✓ Practices done: data transformation and cleaning on “admin”, functionalized the codes and worked on two other data docs, “banking”, “community service”

# Example: Melbourne Housing data exploration and visualization

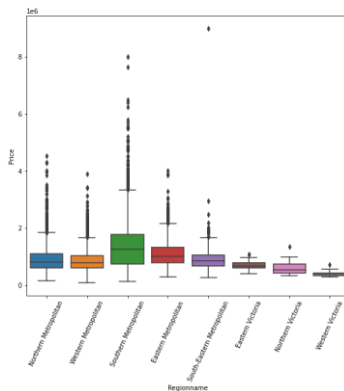
## 1 Univariate analysis

- patterns of each individual feature, categorical or numerical
- common statistics of distributions (mean, median, quantiles etc)
- histogram, charts, box/violin plot, frequency table, density plot etc.



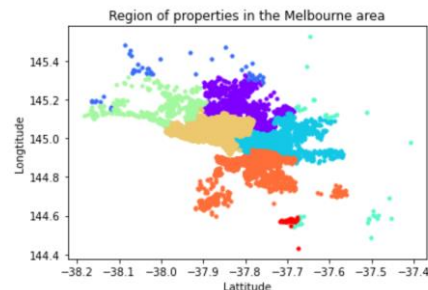
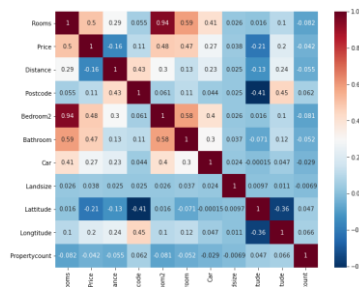
## 2 Bi-/Multi- variate analysis

- relationships/correlations of two or more than two features
- common statistics of distributions (mean, median, quantiles etc)
- Scatter plot, heatmap, facet grid, box/violin plot, pair plot, etc.

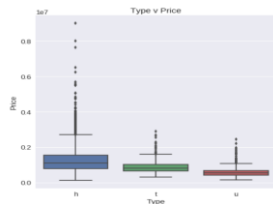


## 3 GIS analysis

- connects data to a map, integrating location data
- More explorations on geospatial analysis by Python

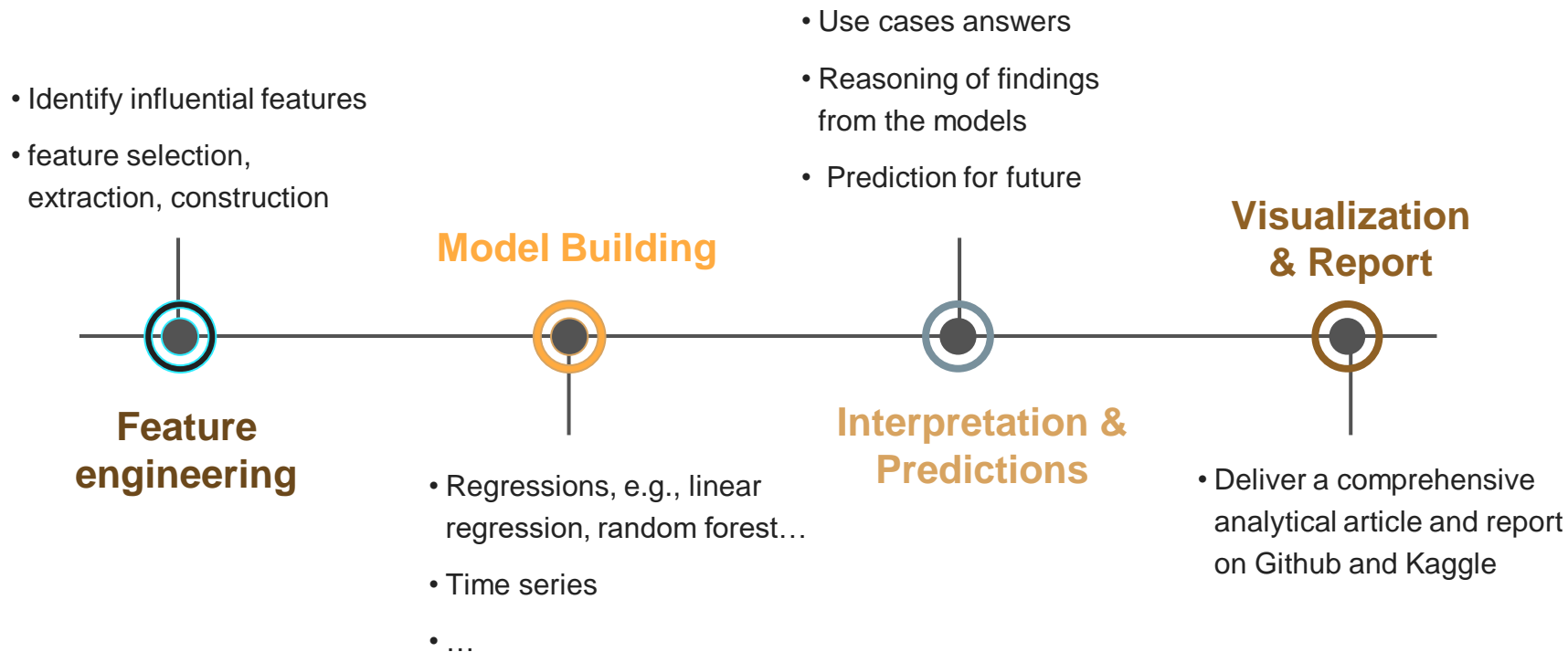


# Example: Melbourne Housing data exploration - some initial findings



- Houses dominate the real estate market in Melbourne and significant differences of prices of different types of properties
- Home prices with different selling methods are relatively the same
- Units have less variant prices according to distance from CBD
- Median prices in the Metropolitan Regions are higher than that of Victoria Region, with Southern Metro having the highest average price
- Price in Southern Metro area seems to be more skewed than that in other areas
- **Positive Correlations:** rooms, bedrooms, bathrooms and prices
- **Weak negative correlation:** price and distance from CBD

# Next plan: Melbourne Housing Data Analysis



**Thank You!**