

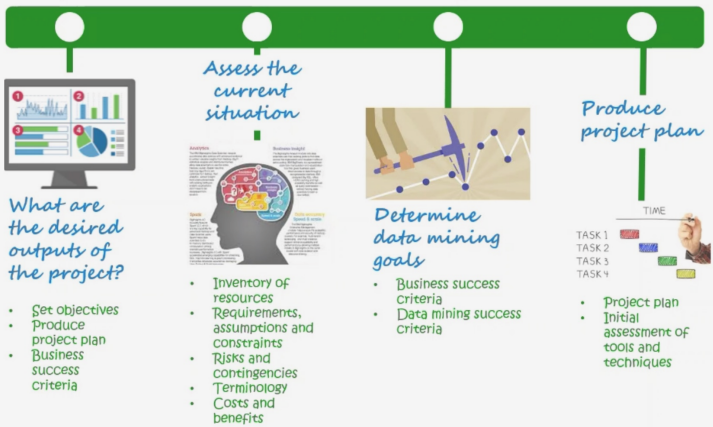


# Agenda

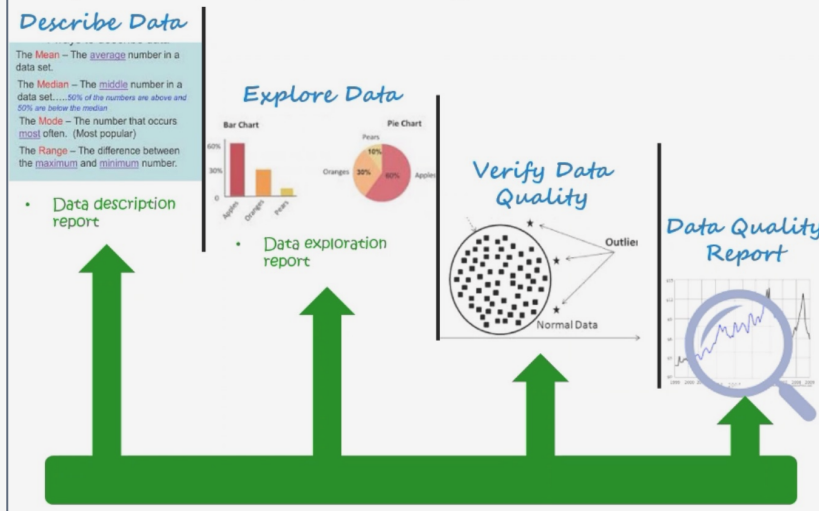
1. Sharing
  - Data Analytics

# Sharing – Data Analytics (1 of 4)

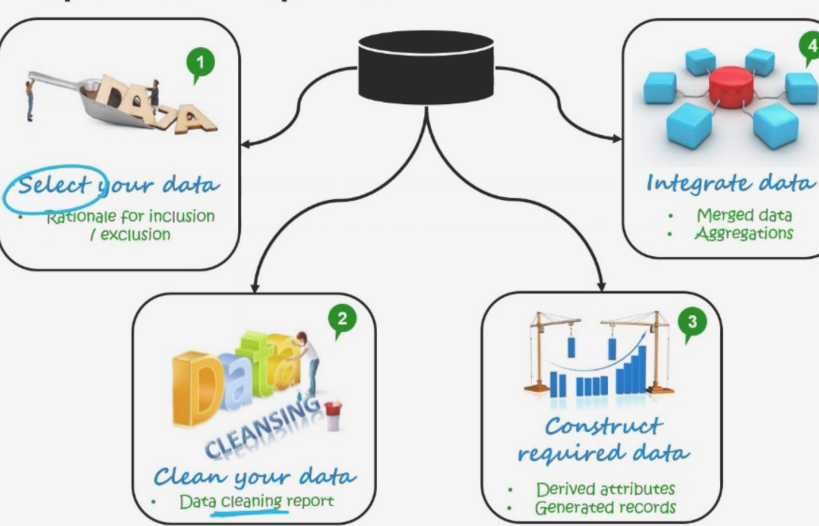
## Step 1: Business Understanding



## Step 2: Data Understanding



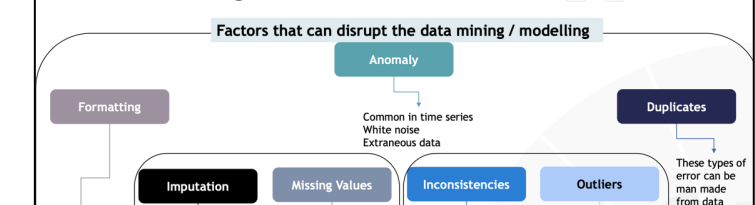
## Step 3: Data Preparation



## Means for Data Visualization:



## Data Cleaning

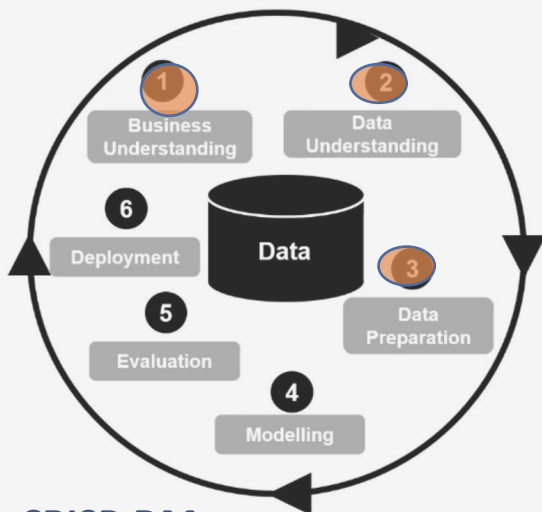


## Data quality checks

1. Count the number of records
2. Add up column if data is complete
3. Check for nulls / blanks
4. Use summary statistics to detect invalid numeric values
5. Dates that are out of range
6. Sort the data and check for top and bottom entries
7. Plot scatter plot using two selected columns to check if data conform to expectation
8. Join to other tables to cross check
9. Check for date and time stamp - no missing gap, no spike (too many records in a time period) and no dip (few or no records)

## CRISP-DM

Cross-InduStrY Processes for Data Mining



# Sharing – Data Analytics (2 of 4)

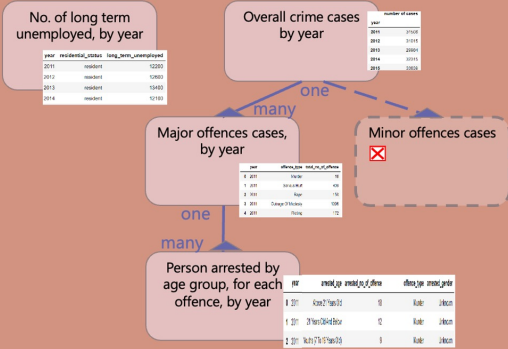
## Data Preparation & Visualization using Python:

### DATA PREPARATION

- Source of data: <<https://data.gov.sg>>

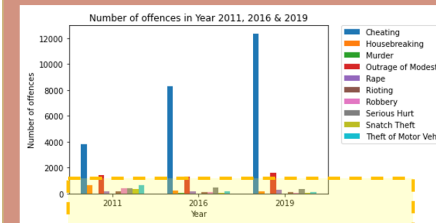
- Data collected:

- ✓ Overall crime cases (Year 2011 – 2020)
- ✓ Crime committed for major offence (Year 2011 – 2020)
- ✓ Person arrested for each major offence, by age group (Year 2011–2019)
- ✓ No. of long-term unemployed residents (Year 1982–2018)



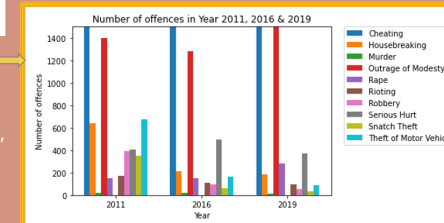
\*\* Plotting of group bar charts, pie charts should be done with year 2011, 2016 and 2019, for clarity of presentation

### DATA ANALYSIS – OVERALL CRIME CASES BY MAJOR OFFENCES TYPE

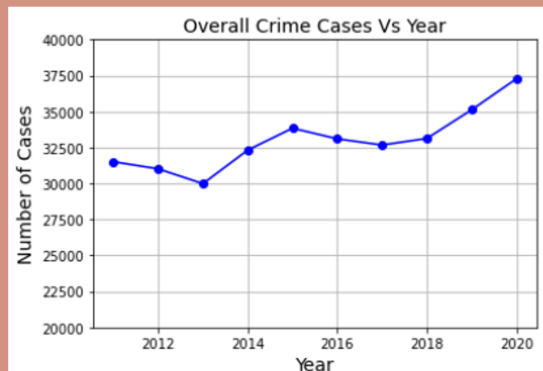


- Breakdown of total crime cases for year 2011, 2016 and 2019 analysed
- Tallest bar is cheating related offences and showed significant increase over the year.

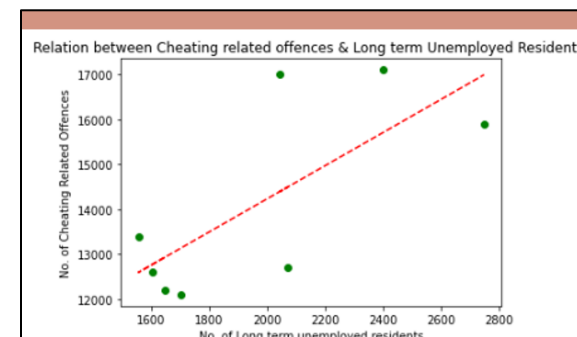
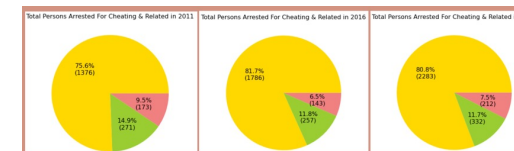
- Similar over the year for murder, rape, serious hurt.
- Significantly decreased for Housebreaking, rioting, robbery, snatch theft and theft of motor vehicle



### DATA ANALYSIS – OVERALL CRIME CASES



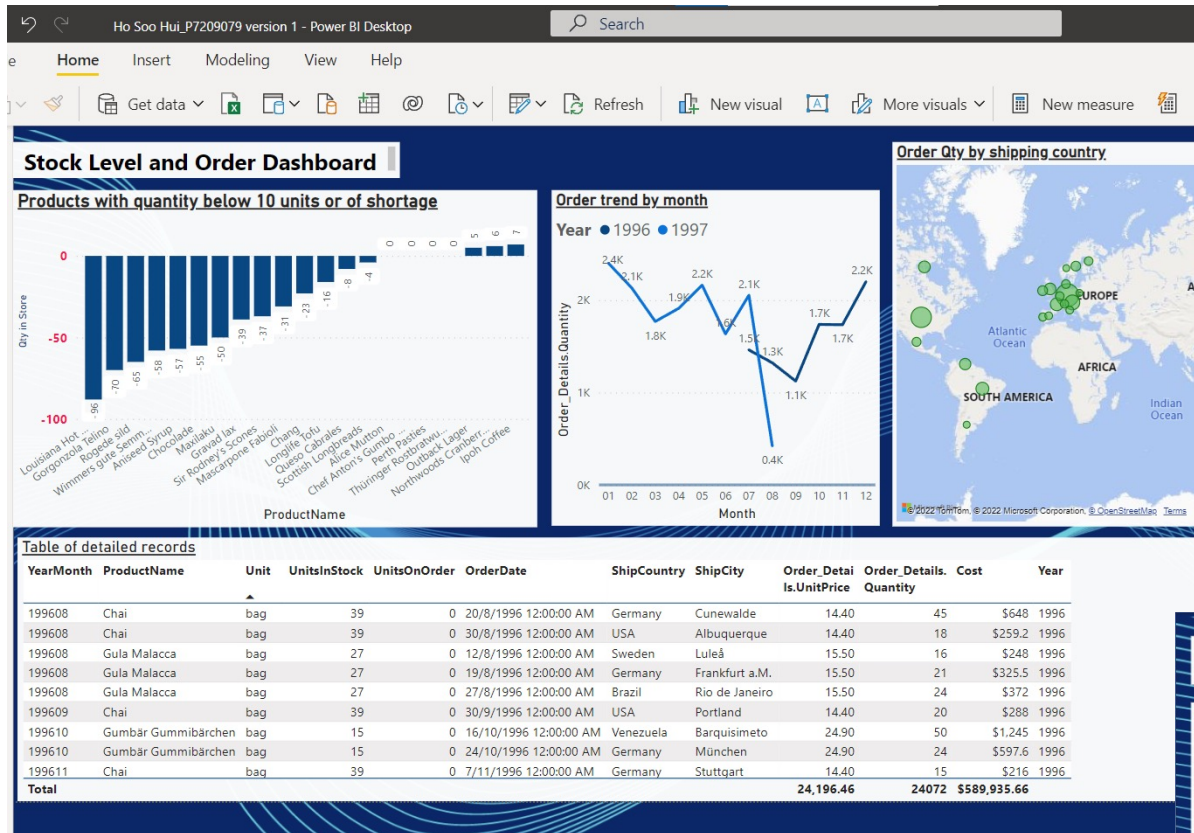
- Overall crime cases increased over the year
- No. of crime cases increased to 37500 cases in 2020 compared to about 31000 cases in 2011



- Weak – Positive relationship between cheating cases (age group above 19 Years Old), and number of long-term unemployed residents for year 2011 to 2018.
- The changes in number of long-term unemployed residents does not cause the increased cheating cases.

# Sharing – Data Analytics (3 of 4)

## Data Preparation & Visualization using Power BI:

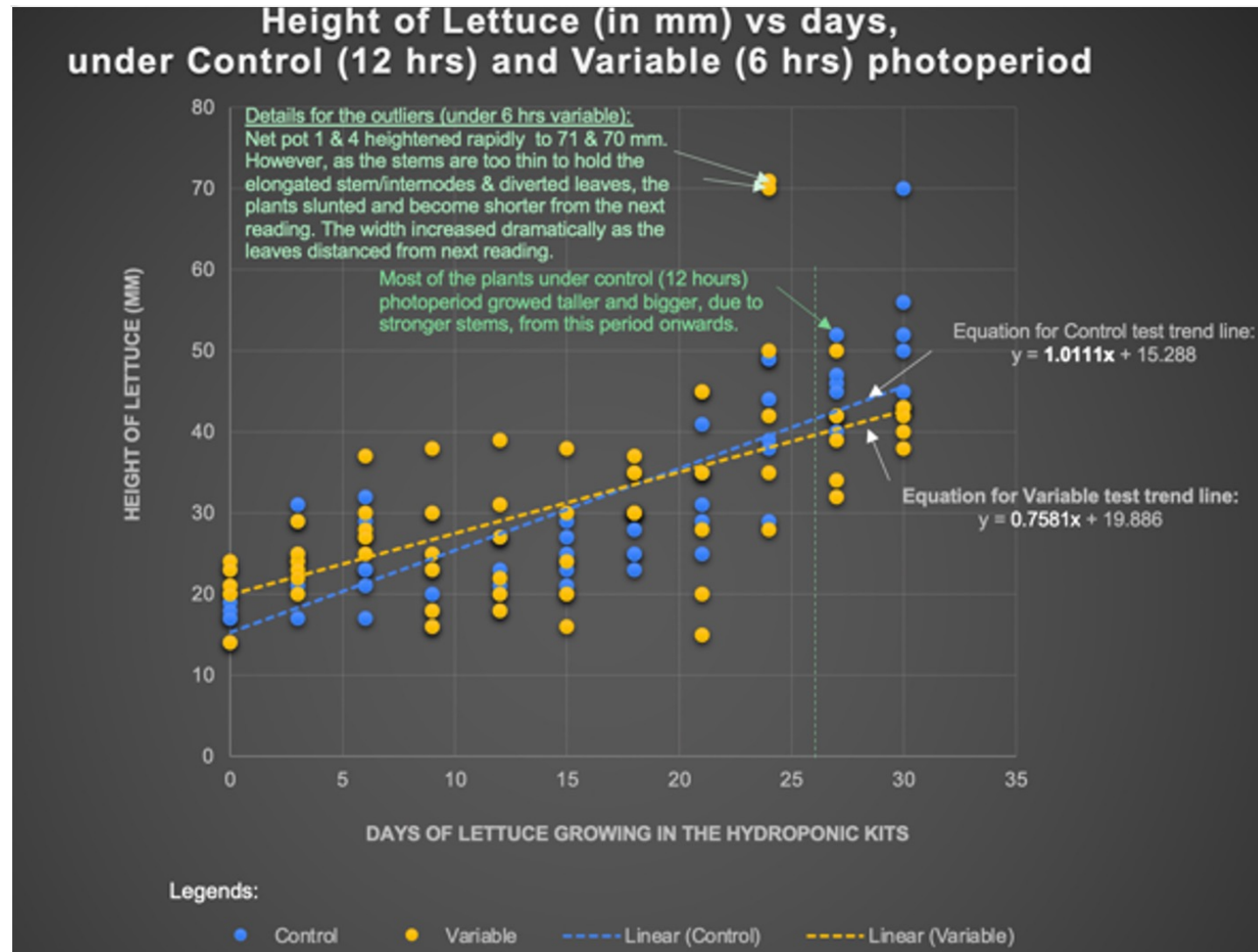


Other tools used include Tableau & Excel

Refer to <https://github.com/Soohui-ho/Data-Science-For-Show>

# Sharing – Data Analytics (4 of 4)

## Data Preparation & Visualization using Excel:





Thank You