



TRAFFIC CONGESTION AND EFFICIENCY IN VICTORIA

WHY AND WHERE DO TRAFFIC CONGESTIONS AND BOTTLENECKS
HAPPEN AND HOW CAN WE IMPROVE ITS EFFICIENCY?

COMP20008 ELEMENTS OF DATA PROCESSING

SEMESTER 1 2016 PROJECT PHASE 3

IVAN KEN WENG CHEE

736901

WHY SHOULD WE CARE ?



GETTING ON TIME

Improve travel times and determine bottlenecks



CATER FOR GROWTH

Amount of traffic in Melbourne rises each year



REDUCE EMISSIONS

Vehicles continue to emit carbon monoxide while idle in traffic jams

RELATED DATASETS

ROAD WIDTH AND NUMBER OF LANES

- **VicRoads Open Data**
- Shows road and shoulder widths and number of traffic lanes on freeways and arterial roads
- http://vicroadsopendata.vicroads.maps.opendata.arcgis.com/datasets/24ccad5c745e4addabfcfb32c400ee83_0

VICTORIAN ROAD TRAFFIC VOLUMES

- **AURIN**
- Contains road traffic volumes for freeways and arterial roads
- <http://data.aurin.org.au/dataset/vic-govt-vicroads-vicroads-evol-mar13-na>

VICROADS SPEED DATA

- **Victorian Government Open Data Repository**
- Records the latest typical hourly speed data in kilometres per hour
- <https://www.data.vic.gov.au/dataset/vicroads-speed-data-by-road-segment>

DATA WRANGLING METHODOLOGIES

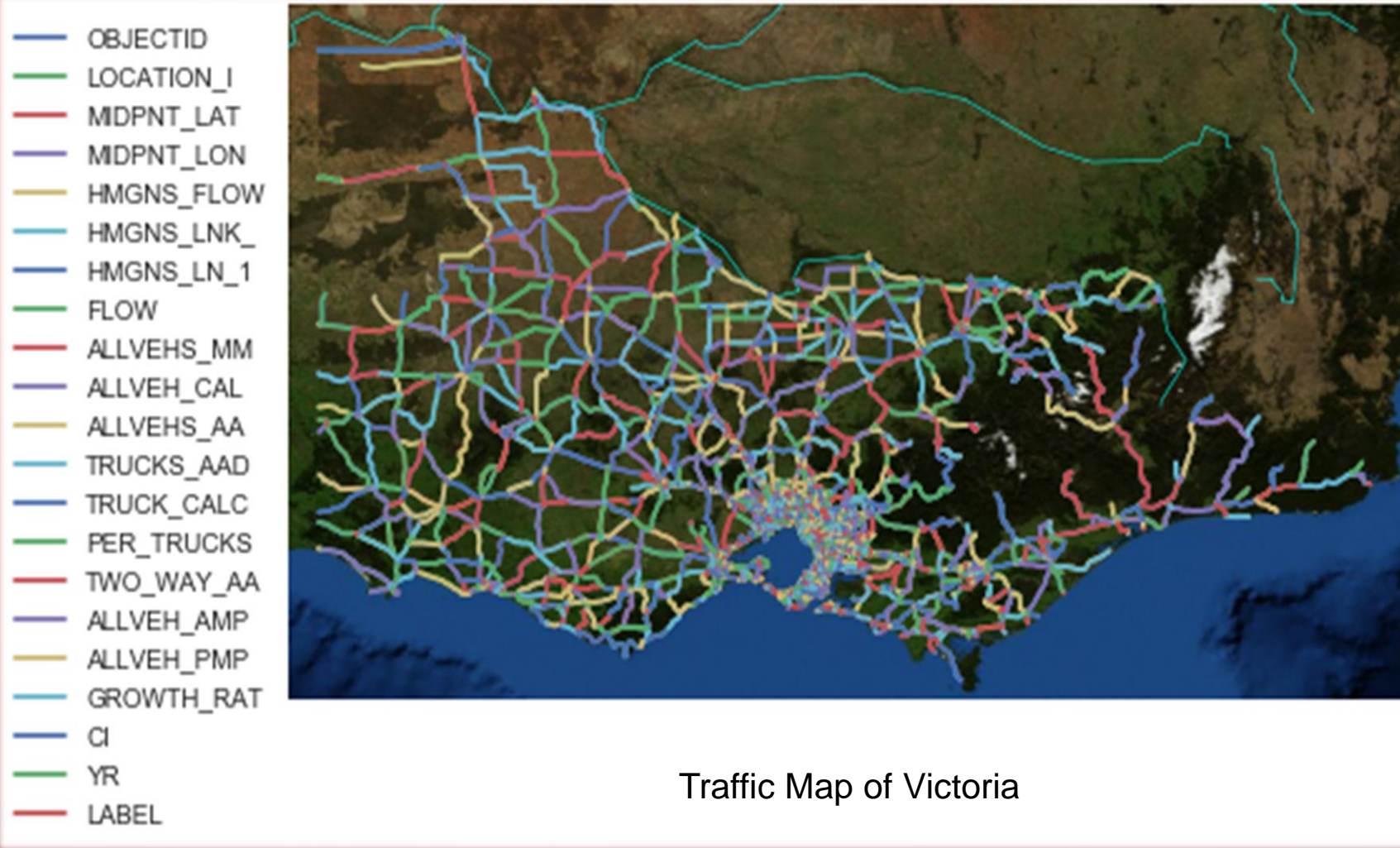
- Extracting
- Munging
- Cleaning
- Aggregation
- Visualisation
- Depositing

The image displays four overlapping Excel workbooks, illustrating the process of data wrangling. The workbooks are:

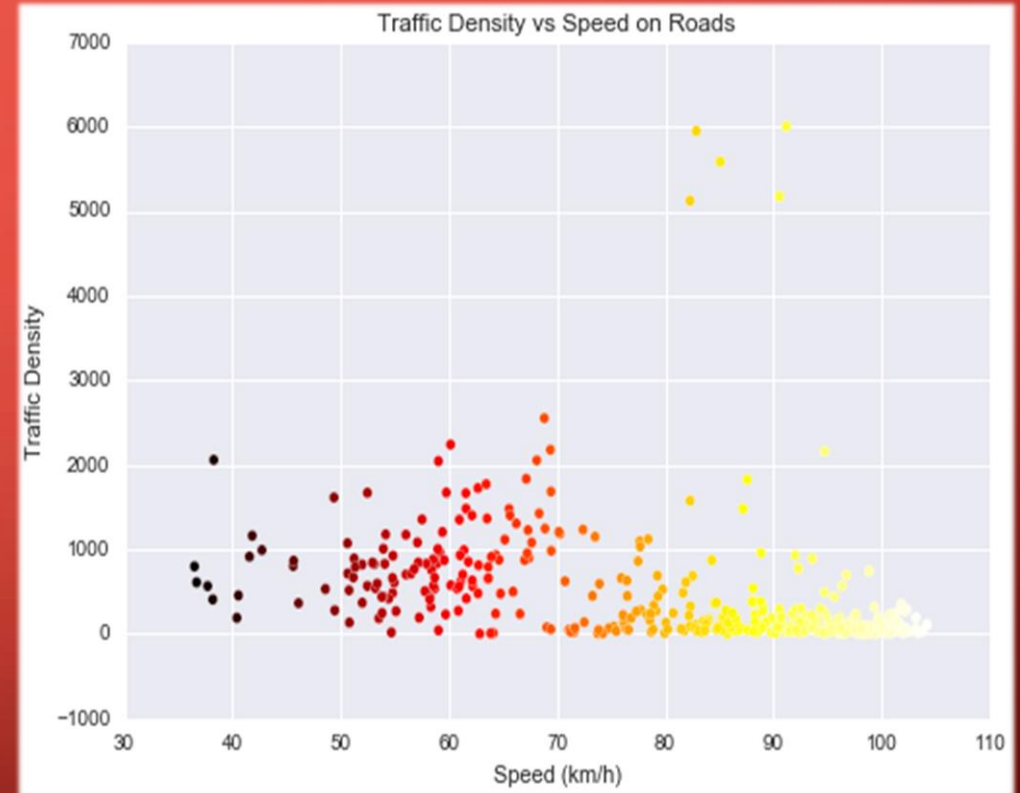
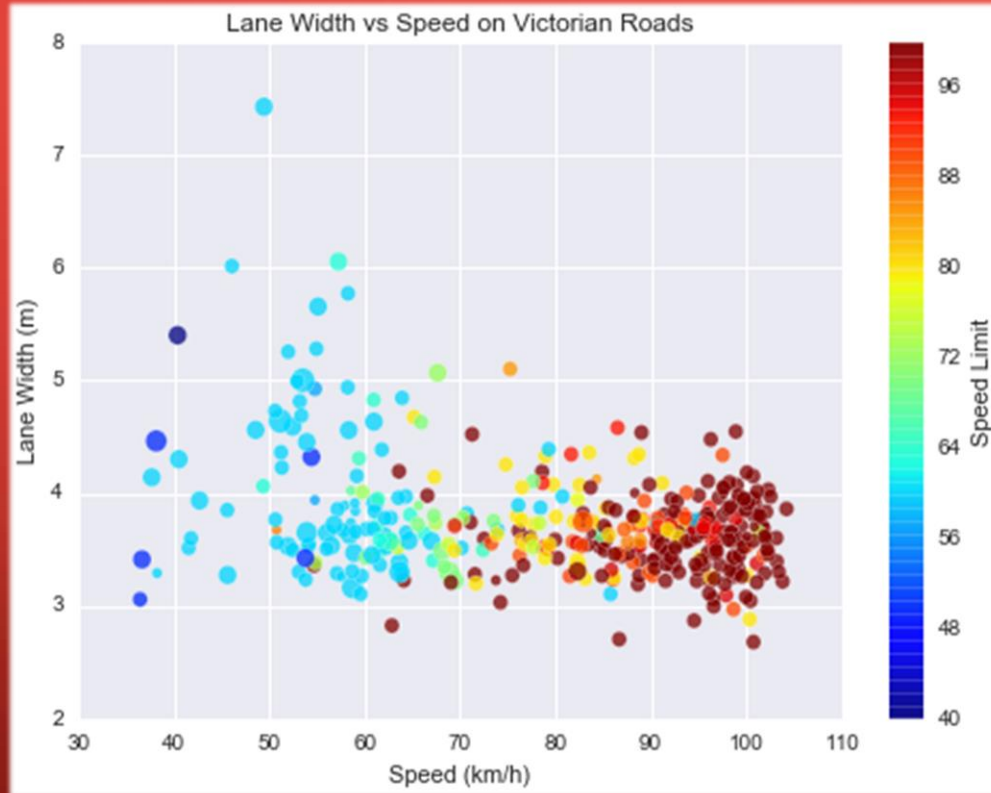
- lanes.csv**: Contains a table with columns A (OBJECTID) and B (ROAD).
- speed.csv**: Contains a table with columns A (NB_TRAFF) and B (NB_L).
- traffic.csv**: Contains a table with columns A (OBJECTID) and B (LOCATION).
- merged.csv**: A larger workbook showing the result of merging data from the other three. It has columns I through S, including LOCATION, MIDPNT_LAT, MIDPNT_LON, HMGNS_F, HMGNS_L, ALLVEHS, ALLVEHS, TRUCKS_AYR, and NB_TRAFF. The data is organized into rows, with some cells containing numerical values and others containing text or formulas.

The merged.csv workbook is the most detailed, showing a large dataset with multiple columns and rows of data. The columns are labeled I through S, and the rows are numbered 1 through 26. The data includes various numerical values, some of which are formatted as currency or percentages. The workbook also shows a status bar at the bottom indicating 'READY' and a zoom level of 100%.

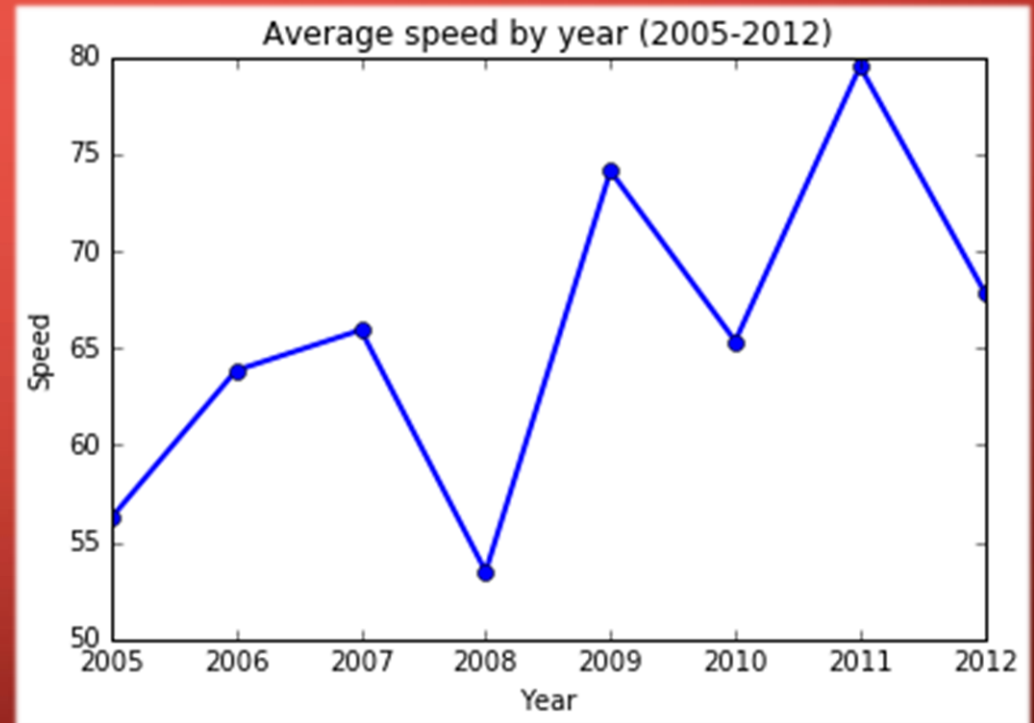
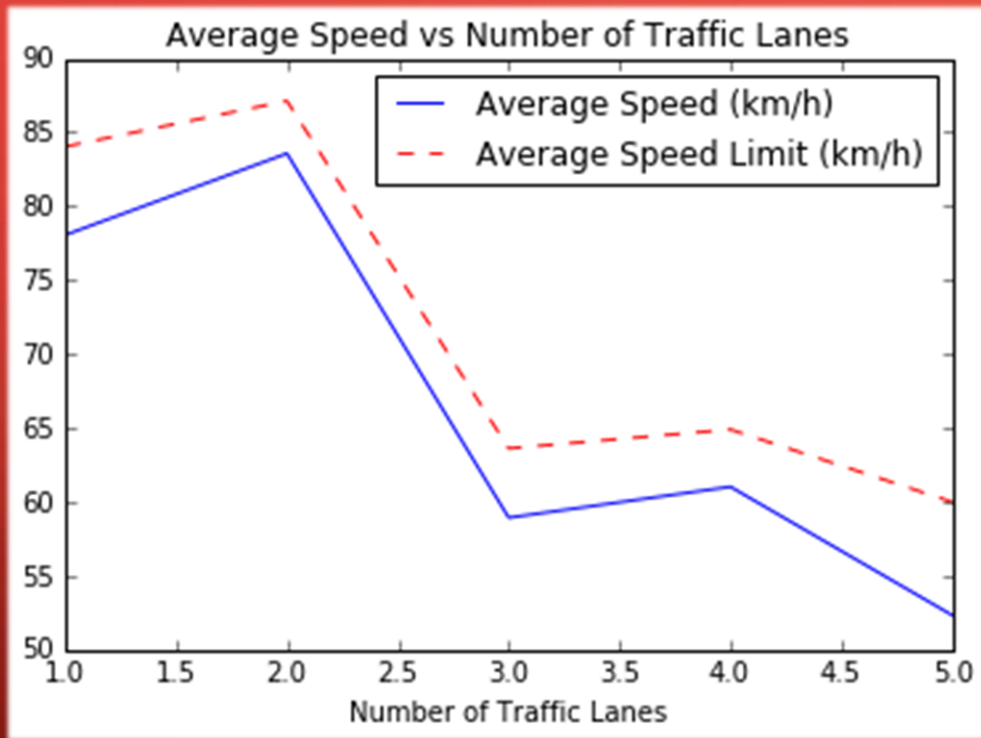
CURRENT FINDINGS



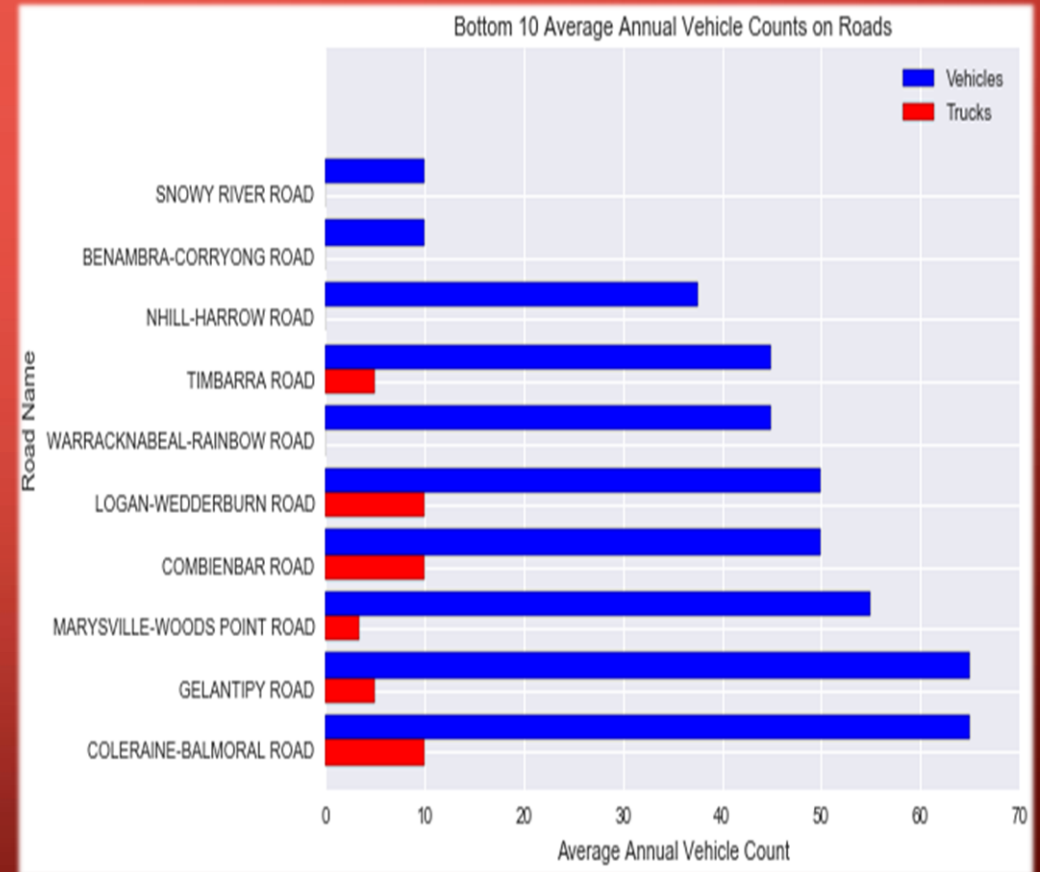
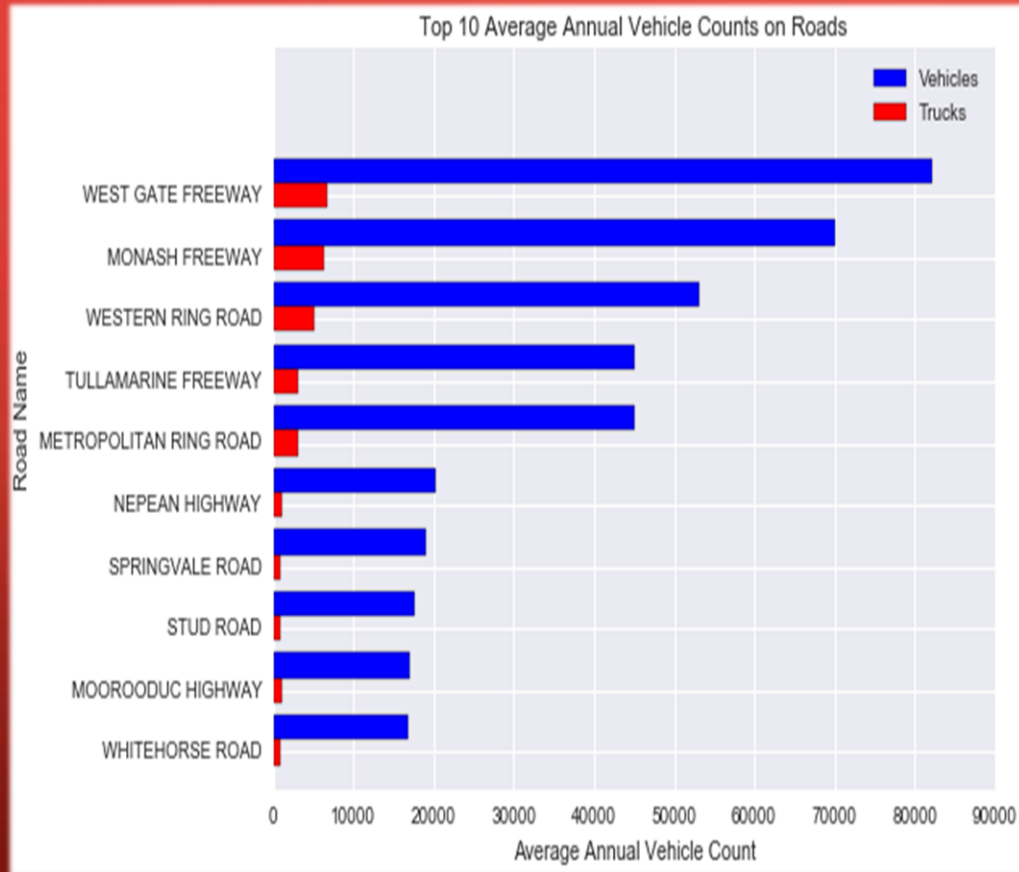
SCATTERPLOTS



LINE PLOTS

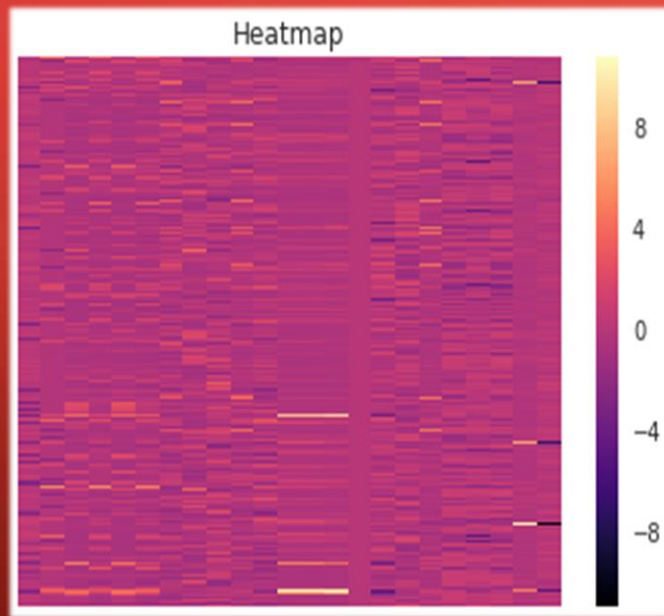


BAR CHARTS

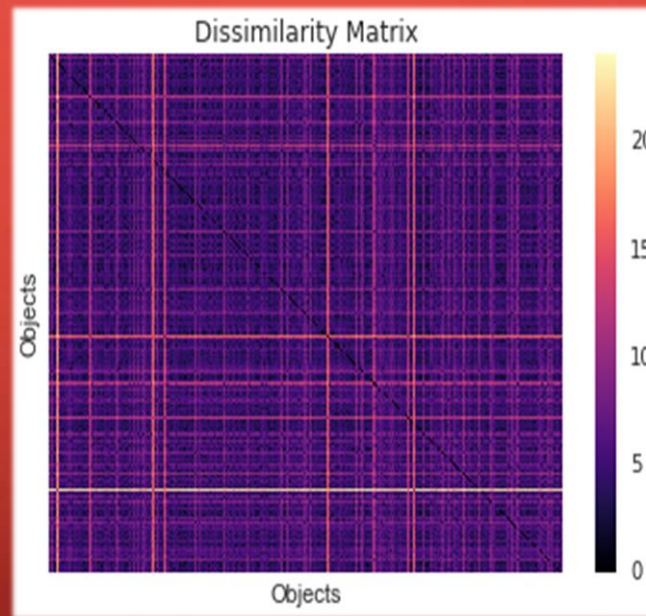


VARIANCE EXPLAINED BY DATA

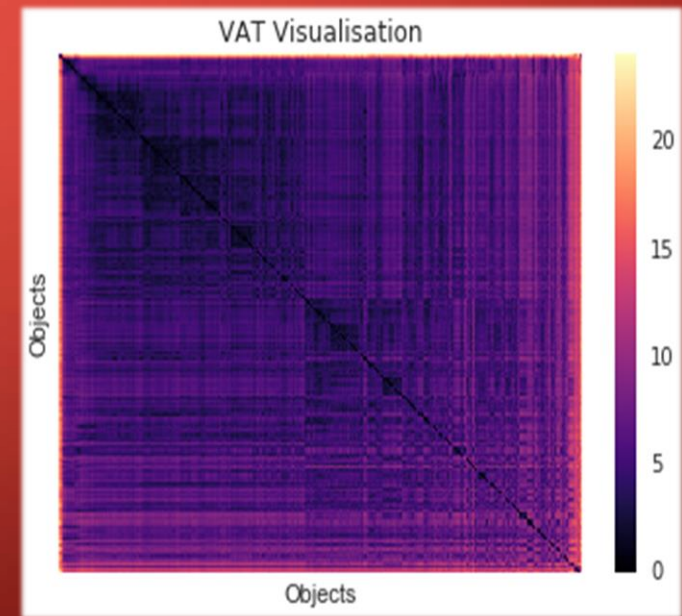
HEATMAP



DISSIMILARITY MATRIX



VAT VISUALISATION



CHALLENGES FACED

- Unconventional formatting of datasets
- Volume of data to be processed
- Different naming conventions in similar columns
- Numerical methods to apply in merging
- Determining non-relevant data to be left out
- Choosing proper visualisation methods

WHAT WOULD I DO DIFFERENTLY

- Better naming conventions
- Choosing more relevant datasets
- Working with json instead of csv