Indonesia Covid Data

Fivetran Ingestion

- Ingested data to Snowflake using google sheet adaptors

DBT Cloud

- Created Project and connected to snowflake 'PLAYGROUNF_GAUTHAM_K" schema
- Used to explore data and deriving columns

Data Modeling

- Created Demographics tables (created as parent-child)
  - Country
  - Island
  - Province
- Created Detailed tables (separated like columns together)
  - Cases_info
  - Recovered_info
  - Deaths_info

Column Featuring

- Area_bucket
  - Bucketed area_km2 column to 3 buckets (small,medium,high)
- Pop_bucket
  - Bucketed population column to 3 buckets (low,medium,high)
- Pop_den_bucket
  - Bucketed population_density column to 3 buckets (low,medium,high)
- Urbanisation_per
  - Percentage of urbanisation per province
  - (Total_rural_villages/total_urban_villages)*100
- Lat_lon
  - Concat(latitude,',',longitude)

*Note: ccolumn bucketing done after checking mean and median

Data observation

- In Source data, country wise daily record is also there. We must take it out for analysis.
- We can compare island/province with that country level data
- We can take rolling average on country wise and compare with state wise

Views Planned

- Province_cases_vw
- Province_death_vw
- Province_recovered_vw
- Cases_wow_vw /mom
- Death_wow_vw/mom
- Recover_wow_vw/mom

Reports Planned

- Trend analysis
- Timeline analysis
  - MoM analysis
  - WoW analysis
  - How weekend affects the trend
- Demographics analysis
  - How area and population density affects the trend
  - How urbanization factor affects the trend
- Comparison analysis
  - How certain province affects country wise