**Data pipeline**

**Technical specification**

**Folder Structure:**

Below is the folder structure used

Datapipeline

export

base

logs

source\_files

staging

preprocess

README.md

Requirements.txt

file\_list.txt

Discount.csv

Config.py

Util.py

Ingestion.py

Preprocess.py

Transformation.py

Export.py

read\_kpi.py

app.py

**Configuration file:**

Config.py will be holding all required configuration to run the project

file\_list.txt will have list of files to be processed

discount.csv will have discount percentage to be processed

**Python files:**

Config.py

* This file will hold configuration file in dictionary format to the data pipeline to use

Util.py

* This file will have common utility to use across the project

Ingestion.py

* This file will input file from source to staging folder

Preprocess.py

* This file will do following task
* 1. Reads the json file
* 2. Converts the json to data frame
* 3. Does the basic data quality
* 4. Generates additional columns
* 5. Filters the data with config value
* 6. Exports the data to folder in CSV format

Transformation.py

* Multiple level of aggregation performed and stored for further processing

Export.py

* This file has function to expose the data into csv format

read\_kpi.py

* This file will have function which will query the final transformed data

app.py

* This is the main entry point for the data pipeline to execute