End to End Data Analytics Project:

Sources, Technologies and Tools

Kaggle Dataset – <https://www.kaggle.com/datasets/ankitbansal06/retail-orders>

Python – Jupyter, Data Cleaning, Processing using Pandas Library

MySQL Database – Install SQL alchemy to connect with MySQL Database

Power BI Tools – Data Analysis and Visualization.

A diagram of data processing

Description automatically generated

Dataset – Download or connect Kaggle Dataset

Requirements –

Jupyter Notebook (Python for Data Analysis) –

Import required libraries such as Pandas and installing sqlalchemy to connect with MySQL Database.

1 Import dataset from Kaggle.

2 Start Processing and Cleaning data, checking Null values

3 Replace ‘Not Available’ and ‘Unknown’ with NaN

4 Convert all column names in lowercase

5 Replace blank space with Underscore ‘\_’ of column names.

6 Create new columns such as Discount, Sales Price and Profit

7 Change format of date in Order Date Column

8 Drop not required columns i.e. List Price, Cost Price, Discount Percent

9 Installing the module which require to connect with python

pip install sqlalchemy

import sqlalchemy as sqc

engine = sqc.create\_engine ('mysql+mysqlconnector://root:Anup%409917@localhost:3306/tasks')

df.to\_sql('orders', con = engine, index=False, if\_exists='append')

10 To append data into database table, create table named ‘Orders’.

Create a table named Orders –

CREATE TABLE orders (

order\_id int PRIMARY KEY, order\_date date,

ship\_mode VARCHAR(20), segment VARCHAR(20), country VARCHAR(20), city VARCHAR(20),

state VARCHAR(20), postal\_code INT, region VARCHAR(20), category VARCHAR(20),

sub\_category VARCHAR(20), product\_id VARCHAR(20), cost\_price INT, list\_price INT, quantity INT,

discount\_percent INT, discount DECIMAL(10,2), sale\_price DECIMAL(10,2), profit DECIMAL(10,2)

);

A screenshot of a computer

Description automatically generated

Check weather data import inside the table.

MySQL Credentials – (Optional)

host='localhost'

user='root'

password='Anup@9917'

database='tasks'