**DATA ANALYSIS BASIC CHEATSHEET**

**Way to use dataset in google colab from google drive**

from google.colab import drive

drive.mount('/content/drive')

# Example path to your dataset:

path = '/content/drive/MyDrive/dataset.csv'

data = pd.read\_csv(path)

data.head()

**1. Importing Essentials**

data = pd.read\_csv('file.csv') # Load CSV

data.head() # First 5 rows

data.tail() # Last 5 rows

data.shape # (rows, columns)

data.columns # Column names

data.info() # Data types + nulls

data.describe() # Summary stats

data.dtypes # Data types

data.index # Index range

2. Load & Inspect Data

data = pd.read\_csv('file.csv') # Load CSV

data.head() # First 5 rows

data.tail() # Last 5 rows

data.shape # (rows, columns)

data.columns # Column names

data.info() # Data types + nulls

data.describe() # Summary stats

data.dtypes # Data types

data.index # Index range

3. Checking Missing/Duplicate values:

data.isnull().sum() # Count missing values per column

data.notnull().sum() # Count non-null values

data.duplicated().sum() # Count duplicate rows

data.drop\_duplicates(inplace=True) # Remove duplicates

data.fillna(0, inplace=True) # Fill NaN with 0 (example)

4. Data Selection & Filtering

data['column'] # Single column

data[['col1', 'col2']] # Multiple columns

data.iloc[0] # Select by index position

data.loc[0] # Select by index label

data.iloc[:, 0:3] # Select columns by index range

data.loc[data['age'] > 30] # Conditional filtering

data.query('age > 30 and gender == "Male"') # Query filtering

5. Sorting and Indexing

data.sort\_values(by='age', ascending=False)

data.reset\_index(drop=True, inplace=True)

data.set\_index('id', inplace=True)

6. Grouping Aggregration and Counting

data['column'].value\_counts() # Frequency count

data['column'].value\_counts(normalize=True) # Percentage

data.groupby('gender')['age'].mean() # Mean age by gender

data.groupby(['gender', 'city']).size() # Group size

data.groupby('department').agg({'salary': ['mean', 'max', 'min']})

7. Cleaning and Encoding

data.rename(columns={'old': 'new'}, inplace=True)

data['gender'].replace({'M': 'Male', 'F': 'Female'}, inplace=True)

pd.get\_dummies(data, columns=['category'], drop\_first=True) # One-hot encoding

8. Visualization Quickies

data.corr() # Correlation matrix

data.cov() # Covariance matrix

data['col'].mean()

data['col'].median()

data['col'].mode()

9.Correlation & Stats

data.corr() # Correlation matrix

data.cov() # Covariance matrix

data['col'].mean()

data['col'].median()

data['col'].mode()

10. Merging joining and concatenation

pd.concat([df1, df2], axis=0) # Stack rows

pd.concat([df1, df2], axis=1) # Combine columns

pd.merge(df1, df2, on='id', how='inner') # SQL-style merge

11. Sampling and Randomization

data.sample(5) # Random 5 rows

data.sample(frac=0.1) # Random 10% sample