**Nicole Gallo**

**D598 – Analytics Programming**

**June 27, 2025**

**QKN1 — QKN1 Task 2: Coding (Attempt 3)**

**A. Create a program in Gitlab using Python to perform the data analysis described in Task1.**

***GitLab link:*** [*https://gitlab.com/wgu-gitlab-environment/student-repos/ngall25/d598-analytics-programming/-/blob/d598\_task2/d598\_dataset.py*](https://gitlab.com/wgu-gitlab-environment/student-repos/ngall25/d598-analytics-programming/-/blob/d598_task2/d598_dataset.py)

import pandas as pd

import numpy as np

# Load dataset

try:

df = pd.read\_excel("D598\_Data\_Set.xlsx", sheet\_name="1-150 V2")

except Exception as e:

raise FileNotFoundError(f"Error loading Excel file: {e}")

# 1. IDENTIFY and REMOVE duplicate rows

duplicates = df[df.duplicated()]

print(f"Duplicate rows found: {len(duplicates)}")

if not duplicates.empty:

df = df.drop\_duplicates()

print("Duplicates removed.")

# 2. GROUP BY 'Business State' and calculate descriptive statistics

required\_columns = [

'Total Revenue', 'Total Long-term Debt', 'Total Equity',

'Total Liabilities', 'Debt to Equity', 'Profit Margin'

]

missing\_cols = [col for col in required\_columns if col not in df.columns]

if missing\_cols:

print(f"Warning: Missing columns for grouping: {missing\_cols}")

df\_by\_state = None

else:

df\_by\_state = df.groupby("Business State")[required\_columns].agg(['mean', 'median', 'min', 'max'])

print("Grouped Statistics by State (Preview):")

print(df\_by\_state.head())

# 3. FILTER for negative Debt to Equity values

if "Debt to Equity" in df.columns:

df\_negative\_dte = df[df["Debt to Equity"] < 0]

print("Businesses with Negative Debt to Equity:")

print(df\_negative\_dte[["Business ID", "Business State", "Debt to Equity"]].head())

else:

print("Warning: 'Debt to Equity' column not found.")

df\_negative\_dte = pd.DataFrame()

# 4. CALCULATE Debt to Income Ratio safely

if "Total Revenue" in df.columns and "Total Long-term Debt" in df.columns:

df["Debt to Income Ratio"] = np.where(

df["Total Revenue"] == 0,

np.nan,

df["Total Long-term Debt"] / df["Total Revenue"]

)

else:

print("Warning: Missing columns for Debt to Income Ratio calculation.")

df["Debt to Income Ratio"] = np.nan

# 5. FINAL DATA EXPORT

df\_combined = df.copy()

df\_combined.to\_excel("Updated\_D598\_Data\_Set\_Final.xlsx", index=False)

print("\nUpdated DataFrame Preview:")

print(df\_combined.head())

**B: Sources** – No external sources were referenced in this task