

Nicole Gallo
D598 – Analytics Programming
June 27, 2025
QKN1 — QKN1 Task 2: Coding

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

```
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