PART 1 — Applied Data Modeling & Reporting (With Kaggle Datasets)

Use Case Example: Customer Churn Reduction

Recommended Kaggle Dataset:

Telco Customer Churn Dataset
🔗 https://www.kaggle.com/blastchar/telco-customer-churn

---

1. Load Dataset & Basic Exploration

```
import pandas as pd

df = pd.read_csv("WA_Fn-UseC_-Telco-Customer-Churn.csv")

print(df.head())
print(df.info())
```

---

2. Check Missing Values

```
print(df.isnull().sum())
```

---

3. Clean Basic Columns

```
df = df.fillna(df.mean()).fillna("Unknown")
```

---

4. Correlation Analysis

```
correlation = df.corr(numeric_only=True)
print(correlation["Churn"].sort_values(ascending=False))
```

---

5. Simple Rule-Based Churn Model

```
df["Risk_Score"] = 0
df["Risk_Score"] += (df["MonthlyCharges"] > 70).astype(int)
df["Risk_Score"] += (df["tenure"] < 6).astype(int)

df["Predicted_Churn"] = (df["Risk_Score"] >= 2).astype(int)

print(df[["MonthlyCharges","tenure","Risk_Score","Predicted_Churn"]].head()
)
```

---

6. Reporting Summary

```
print("Total Customers:", len(df))
print("Actual Churn:", df["Churn"].mean())
print("Predicted Churn:", df["Predicted_Churn"].mean())
```

PART 2 — Applied Data Modeling & Reporting (With Kaggle Dataset)

Use Case Example: Operations Process Optimization

Recommended Kaggle Dataset:

📌 **UCI Online Retail Dataset (Transaction + Operations)**
🔗 https://www.kaggle.com/carrie1/ecommerce-data

---

1. Load Dataset

```
df = pd.read_csv("data.csv", encoding="ISO-8859-1")
print(df.head())
```

---

2. Compute Processing Time (Example for Operations Data)

(If dataset doesn't have timestamps → use synthetic columns in class.)

```
df["Processing_Time"] = df["Quantity"] * 2   # fake logic for training only
df["Delay"] = df["Processing_Time"] - 5
```

---

3. Identify Top Problem Areas

```
print(df.sort_values(by="Delay", ascending=False).head(10))
```

---

4. Operations Insights

```
print("Average Delay:", df["Delay"].mean())
print(df.groupby("Country")["Delay"].mean().sort_values(ascending=False))
```

CAPSTONE PROJECT (With Kaggle Dataset)

Title:

Customer Churn Prediction + Reporting Dashboard

Kaggle Dataset (Recommended):

Telco Churn Dataset
https://www.kaggle.com/blastchar/telco-customer-churn

---

1. Import & Clean Data

```
import pandas as pd

df = pd.read_csv("WA_Fn-UseC_-Telco-Customer-Churn.csv")
df = df.fillna(df.mean()).fillna("Unknown")
```

---

2. Create Simple Features

```
df["HighCharge"] = (df["MonthlyCharges"] > 70).astype(int)
df["ShortTenure"] = (df["tenure"] < 6).astype(int)
```

---

3. Rule-Based Prediction Model

```
df["RiskScore"] = df["HighCharge"] + df["ShortTenure"]
df["Prediction"] = (df["RiskScore"] >= 1).astype(int)
```

4.Summary Report

```
report = {
    "Total Customers": len(df),
    "Predicted Churn Rate": df["Prediction"].mean(),
    "High Risk Customers": df[df["RiskScore"] >= 1].shape[0]
}
print(report)
```

---

5. Optional Visualization

```
df["Prediction"].value_counts().plot(kind="bar")
```