Supply Chain Dataset Analysis Project Guide

Week 1: Data Preprocessing (Data Engineer)

```
Goal: Prepare clean, structured data.
Tools: Python (pandas, SQL), Jupyter Notebook

Steps:
1. Load Data
   ```python
   import pandas as pd
   df = pd.read_csv("supply_chain_data.csv")
   ```

2. Inspect Dataset
   ```python
   df.info()
   df.describe()
   df.isnull().sum()
   ```

3. Clean Columns
   ```python
   df['Inspection results'] = df['Inspection results'].str.strip().str.title()
   ```

4. Feature Engineering
   ```python
   df['Cost per unit'] = df['Manufacturing costs'] / df['Production volumes']
   ```

5. Export Clean Data
   ```python
   df.to_csv("clean_supply_chain_data.csv", index=False)
   ```

Deliverables:
- Clean CSV dataset
- Preprocessing notebook
```

Week 2: Analysis Questions (Business Analyst)

```
Goal: Ask and answer supply chain analysis questions.
Tools: SQL / Python (pandas, matplotlib)

Example Questions:
1. Which transportation mode is fastest vs most expensive?
   ```python
   df.groupby("Transportation modes")[["Lead time","Manufacturing costs"]].mean()
   ```

2. Do higher defect rates increase manufacturing cost?
   ```python
   df.plot.scatter(x="Defect rates", y="Manufacturing costs")
   ```

3. Which routes are most cost-efficient?
   ```python
   df.groupby("Routes")["Manufacturing costs"].mean().sort_values()
   ```

SQL Example:
   ```sql
   SELECT Routes, AVG(Manufacturing_costs) AS avg_cost
   FROM supply_chain
   GROUP BY Routes
   ORDER BY avg_cost;
   ```

Deliverables:
```

```
- List of analysis questions
- Tables/graphs answering them
```

Week 3: Forecasting (Data Scientist)

```
Goal: Build forecasting models.
Tools: Python (scikit-learn, pandas, matplotlib)

Steps:
1. Forecast Manufacturing Costs from Production Volumes
```

```python
from sklearn.linear_model import LinearRegression
X = df[['Production volumes']]
y = df['Manufacturing costs']
model = LinearRegression().fit(X, y)
```

```
2. Predict Lead Time Based on Transportation & Route
```

```python
df = pd.get_dummies(df, columns=['Transportation modes','Routes'], drop_first=True)
X = df.drop("Lead time", axis=1)
y = df["Lead time"]
```

```
3. Forecast Defect Rate Trend (Regression or Time Series)

Deliverables:
- Forecasting notebook
- Prediction graphs
```

Week 4: Visualization & Storytelling (Visualization Specialist)

```
Goal: Build Tableau dashboard & final presentation.
Tools: Tableau, PowerPoint, SQL (if needed)

Steps:
1. Import Clean Dataset into Tableau.
2. Create KPIs (Avg Lead Time, Cost per Unit, Defect Rate).
3. Build Visuals:
   - Bar chart: Transportation mode vs cost & lead time
   - Line chart: Production volume vs cost trend
   - Map: Costs by city
   - Forecast chart: Predicted defect rate

4. Dashboard Layout:
   - Top: KPI Cards
   - Left: Transportation & Route Analysis
   - Right: Forecast Plots
   - Bottom: Cost vs Defect Trends

5. Presentation: Problem → Dataset → Insights → Forecasts → Recommendations

Deliverables:
- Tableau Dashboard
- Final Presentation
```

Team Roles Summary

```
- Data Engineer: Cleans and preprocesses dataset, delivers ready-to-use data.
- Business Analyst: Poses supply chain questions, runs SQL/pandas analysis, delivers insights.
- Data Scientist: Builds predictive models, delivers forecast plots.
- Visualization Specialist: Creates Tableau dashboard & final presentation.
```