TASK 2: Data Preparation and Integration

```
import pandas as pd


file_path = r"C:\Users\91952\Documents\ACADEMICS\Programming 1x\
Projects\Axion Assessment\Data for Task 2.xlsx"


df1 = pd.read_excel(file_path, sheet_name=0)
df2 = pd.read_excel(file_path, sheet_name=1)

df1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 500 entries, 0 to 499
Data columns (total 15 columns):
 #   Column                                  Non-Null Count  Dtype

---  ------                                  --------------  -----

 0   Primary Key                             500 non-null    object

 1   Order No                                500 non-null    object

 2   Segment Number                          500 non-null    int64

 3   Order Date                              500 non-null
datetime64[ns]
 4   Manufacturer                            500 non-null    object

 5   Model                                   500 non-null    object

 6   Product Category                        500 non-null    object

 7   Model Year                              500 non-null    int64

 8   Serial Number                           500 non-null    object

 9   Meter 1 Reading                         500 non-null    float64

 10  Complaint                               500 non-null    object

 11  Cause                                   206 non-null    object

 12  Correction                              475 non-null    object

 13  Failure Condition - Failure Component  500 non-null    object

 14  Fix Condition - Fix Component           500 non-null    object
```

```
dtypes: datetime64[ns](1), float64(1), int64(2), object(11)
memory usage: 58.7+ KB
```

```
df2.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 500 entries, 0 to 499
Data columns (total 13 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   Primary Key        500 non-null    object
 1   Order No           500 non-null    object
 2   Segment Number     500 non-null    int64
 3   Coverage           81 non-null     object
 4   Qty                500 non-null    int64
 5   Part Manufacturer  500 non-null    object
 6   Part Number        500 non-null    object
 7   Part Description   500 non-null    object
 8   Revenue            500 non-null    float64
 9   Cost               500 non-null    object
 10  Invoice Date       500 non-null    int64
 11  Actual Hours       482 non-null    float64
 12  Segment Total $    500 non-null    object
dtypes: float64(2), int64(3), object(8)
memory usage: 50.9+ KB
```

```
df1.head(2)
```

```
   Primary Key   Order No  Segment Number Order Date Manufacturer
Model  \
0  SO0005588-1  SO0005588               1 2022-04-30       PASEIH
6780
1  SO0005907-1  SO0005907               1 2022-04-30       PASEIH
6780

  Product Category  Model Year Serial Number  Meter 1 Reading  \
0             APPL           0     YFT042399        2531.0999
1             APPL           0     YFT042399        2531.0999

                                          Complaint Cause  \
0  No cab heat, temp gauge dont get to operating ...   NaN
1  No cab heat, temp gauge dont get to operating ...   NaN

                                        Correction  \
0  Als ich das Gerät in die Werkstatt fuhr, stieg...
1  Als ich das Gerät in die Werkstatt fuhr, stieg...

  Failure Condition - Failure Component   Fix Condition - Fix
Component
0  No Heat - Cab, Not Achieving - Gauge  No Component Mentioned -
```

```
Added
1  No Heat - Cab, Not Achieving - Gauge  No Component Mentioned -
Added

df2.head(3)

   Primary Key   Order No  Segment Number           Coverage  Qty  \
0  SO0005588-1  SO0005588               1  mike 102-305-1811   37
1  SO0005907-1  SO0005907               1  mike 102-305-1811    1
2  SO0006100-1  SO0006100               1  mike 102-305-1811    3

  Part Manufacturer          Part Number   Part Description   Revenue
\
0              PASE  042094R9-Q     PASE   NO.1-15W40 CJ4QT  127.2799

1              PASE    25505353     PASE              FLUID   30.0000

2              PASE    25500540     PASE  ACTIFUL OT PREMIX  126.0000


       Cost  Invoice Date  Actual Hours Segment Total $
0  96.1999$         44698        6.3798      1048.3596$
1    22.68$         44698        6.3798      1048.3596$
2     78.3$         44698        6.3798      1048.3596$
```

Both df1 and df2 have the column `"Primary Key"`. We'll use `"Primary Key"` as the join key.

Check for Nulls:

- df1:

```
print("Nulls in df1:")
print(df1.isnull().sum(), "\n")

Nulls in df1:
Primary Key                                0
Order No                                   0
Segment Number                             0
Order Date                                 0
Manufacturer                               0
Model                                      0
Product Category                           0
Model Year                                 0
Serial Number                              0
Meter 1 Reading                            0
Complaint                                  0
Cause                                    294
Correction                                25
Failure Condition - Failure Component      0
Fix Condition - Fix Component              0
```

```
dtype: int64
```

- df2:

```
print("Nulls in df2:")
print(df2.isnull().sum(), "\n")
```

```
Nulls in df2:
Primary Key            0
Order No               0
Segment Number         0
Coverage             419
Qty                    0
Part Manufacturer      0
Part Number            0
Part Description       0
Revenue                0
Cost                   0
Invoice Date           0
Actual Hours          18
Segment Total $        0
dtype: int64
```

Drop duplicates if any

```
df1.drop_duplicates(subset="Primary Key", inplace=True)
df2.drop_duplicates(subset="Primary Key", inplace=True)
```

Standardize column names (strip whitespace)

```
df1.columns = df1.columns.str.strip()
df2.columns = df2.columns.str.strip()

print("\nData types:")
print(df1.dtypes, "\n")
print(df2.dtypes)
```

```
Data types:
Primary Key                                 object
Order No                                    object
Segment Number                               int64
Order Date                          datetime64[ns]
Manufacturer                                object
Model                                       object
Product Category                            object
Model Year                                   int64
Serial Number                               object
```

```
Meter 1 Reading                                    float64
Complaint                                           object
Cause                                               object
Correction                                          object
Failure Condition - Failure Component               object
Fix Condition - Fix Component                       object
dtype: object
```

```
Primary Key           object
Order No              object
Segment Number         int64
Coverage              object
Qty                    int64
Part Manufacturer     object
Part Number           object
Part Description      object
Revenue              float64
Cost                  object
Invoice Date           int64
Actual Hours         float64
Segment Total $       object
dtype: object
```

Convert cost and segment total to float

```
df2['Cost'] = pd.to_numeric(df2['Cost'], errors='coerce')
df2['Segment Total $'] = pd.to_numeric(df2['Segment Total $'],
errors='coerce')
```

Replace null 'Coverage' with 'Unknown'

```
df2['Coverage'] = df2['Coverage'].fillna('Unknown')

df2[['Coverage']].isnull().sum()
```

```
Coverage    0
dtype: int64
```

Fill 'Cause' and 'Correction' in df1 with 'Not Mentioned'

```
df1['Cause'] = df1['Cause'].fillna('Not Mentioned')
df1['Correction'] = df1['Correction'].fillna('Not Mentioned')

df1[['Cause', 'Correction']].isnull().sum()
```

```
Cause         0
Correction    0
dtype: int64
```

Merge on Primary Key using LEFT JOIN to preserve all rows from df1

```
merged_df = pd.merge(df1, df2, on="Primary Key", how="left")

merged_df.head(3)
```

```
   Primary Key Order No_x  Segment Number_x Order Date Manufacturer
Model  \
0  SO0005588-1  SO0005588                 1 2022-04-30       PASEIH
6780
1  SO0005907-1  SO0005907                 1 2022-04-30       PASEIH
6780
2  SO0006100-1  SO0006100                 1 2022-04-30       PASEIH
6780

  Product Category  Model Year Serial Number  Meter 1 Reading  ...  \
0             APPL           0     YFT042399        2531.0999  ...
1             APPL           0     YFT042399        2531.0999  ...
2             APPL           0     YFT042399        2531.0999  ...

            Coverage   Qty Part Manufacturer          Part Number  \
0  mike 102-305-1811  37.0              PASE  042094R9-Q     PASE
1  mike 102-305-1811   1.0              PASE    25505353     PASE
2  mike 102-305-1811   3.0              PASE    25500540     PASE

    Part Description   Revenue  Cost Invoice Date  Actual Hours  \
0   NO.1-15W40 CJ4QT  127.2799   NaN      44698.0        6.3798
1              FLUID   30.0000   NaN      44698.0        6.3798
2  ACTIFUL OT PREMIX  126.0000   NaN      44698.0        6.3798

  Segment Total $
0             NaN
1             NaN
2             NaN

[3 rows x 27 columns]
```

```
merged_df.isnull().sum()
```

```
Primary Key                                0
Order No_x                                 0
Segment Number_x                           0
Order Date                                 0
Manufacturer                               0
Model                                      0
Product Category                           0
Model Year                                 0
Serial Number                              0
Meter 1 Reading                            0
Complaint                                  0
Cause                                      0
```

```
Correction                                    0
Failure Condition - Failure Component         0
Fix Condition - Fix Component                 0
Order No_y                                    5
Segment Number_y                              5
Coverage                                      5
Qty                                           5
Part Manufacturer                             5
Part Number                                   5
Part Description                              5
Revenue                                       5
Cost                                        500
Invoice Date                                  5
Actual Hours                                 23
Segment Total $                             500
dtype: int64
```

TASK 3.1 – Trend Analysis (Visual + Insights)

```
import matplotlib.pyplot as plt
import seaborn as sns
```

Make sure Order Date is datetime

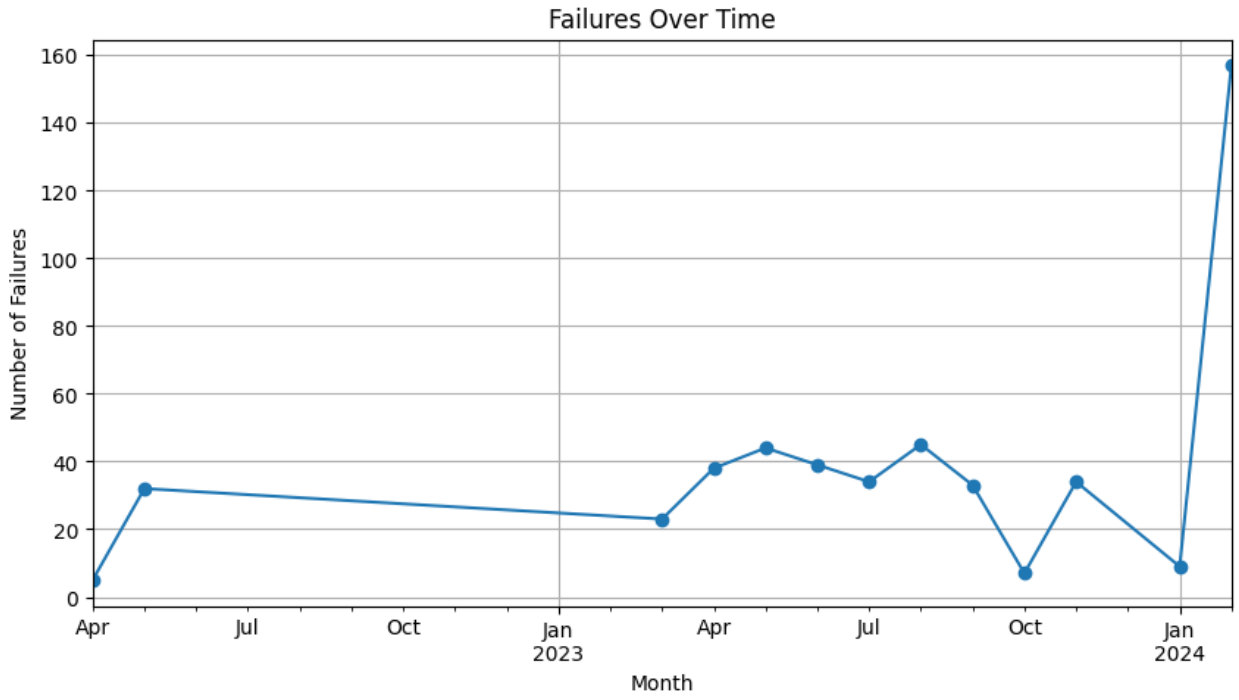
```
merged_df['Order Date'] = pd.to_datetime(merged_df['Order Date'])
```

Count by Month

```
trend = merged_df.groupby(merged_df['Order
Date'].dt.to_period('M')).size()
```

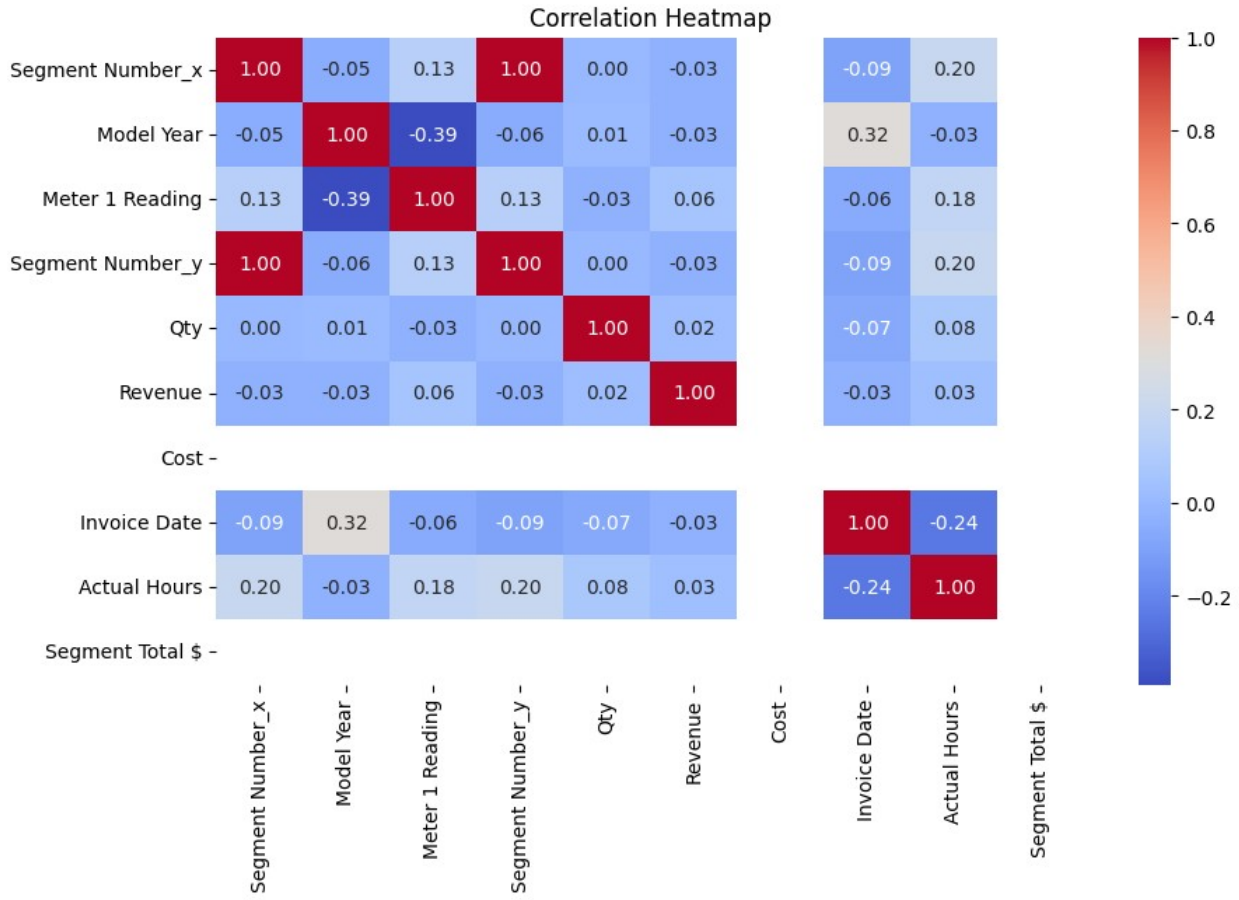
Plot

```
trend.plot(kind='line', marker='o', title="Failures Over Time",
figsize=(10,5))
plt.xlabel("Month")
plt.ylabel("Number of Failures")
plt.grid()
plt.show()
```



Heatmap for numeric columns

```
plt.figure(figsize=(10,6))
sns.heatmap(merged_df.corr(numeric_only=True), annot=True,
cmap='coolwarm', fmt=".2f")
plt.title("Correlation Heatmap")
plt.show()
```



Breakdown:

- ***Correlation Coefficient:*** The numbers within each cell (e.g., 1.00, -0.05, 0.13) are the Pearson correlation coefficients, ranging from -1 to +1.
  - +1.00: Perfect positive correlation (as one variable increases, the other increases proportionally).
  - -1.00: Perfect negative correlation (as one variable increases, the other decreases proportionally).
  - 0.00: No linear correlation.
  - Values closer to +1 or -1: Stronger correlation.
  - Values closer to 0: Weaker correlation.
- ***Color Scale (Coolwarm):*** The color bar on the right indicates the range of correlation values.
  - Red (warm colors): Indicate positive correlations. Darker red means stronger positive correlation.
  - Blue (cool colors): Indicate negative correlations. Darker blue means stronger negative correlation.
  - White/Light colors: Indicate correlations close to zero (weak or no linear correlation).

TASK 3.2 – Root Cause Identification

```
merged_df['Failure Condition - Failure Component'].unique()
```

```
array(['No Heat - Cab, Not Achieving - Gauge',
       'Not Charging - Alternator', 'Faulty - Fan',
       'Oil Loss - Not Mentioned',
       'Unavailable - Suspension, Unavailable - Axle',
       'Not Mentioned - No Component Mentioned',
       'Missed - No Component Mentioned, Missed - No Component
Mentioned',
       'Leak - Coupler', 'Missed - Boom', 'Error Code - Nozzle',
       'Not Opening - Door',
       'Loose - Clip, Loose - Duct, Not Locked - Bulkhead',
       'Crushed - Compressor', 'Oil Leak - Machine',
       'Not Used - Mirror, Not Mentioned - Valve, Not Mentioned - NCV,
Not Mentioned - Nozzle, Not Mentioned - Cap',
       'Oil Leak - Gear Pump, Oil Leak - Port, Water Leak - Air Line,
Rubbed - Shield, Damaged - Hydraulic System',
       'Hit - Tank, Hit - Rail, Not Cradling - Boom', 'Leak - Hose',
       'Oil Leak - No Component Mentioned, Damaged - O Ring, Protruded
- O Ring',
       'Broken - Harness', 'Leak - Tank',
       'Not Mentioned - Light, Incorrect - No Component Mentioned, Not
Settling - No Component Mentioned, Not Settling - Boom, Low - Boom,
Hard To Set - Boom, Leak - Boom',
       'Not Installed - Nozzle', 'Oil Leak - Steering',
       'Open - Fuel System', 'Oil Leak - Tire', 'Leak - Cooler',
       'Frozen - No Component Mentioned', 'Leak - Boom',
       'Hit - Nozzle, Damaged - No Component Mentioned',
       'Not Mentioned - Engine, Not Mentioned - Fuel Filter, Not
Mentioned - Filter, Not Mentioned - Hydraulic System',
       'Leak - Hydraulic Tank',
       'Blocked - Hub, Not Mentioned - Hydraulic Filter, Not Mentioned
- Filter, Not Mentioned - Engine, Not Mentioned - Filter',
       'Not Mentioned - Not Mentioned', 'Leak - Nozzle',
       'Error Code - Sensor', 'Inoperative - Boom', 'Error Code -
Module',
       'Inconsistent Folding - Boom, Hit - Cradle', 'Smashed -
Camera',
       'Not Mentioned - Machine', 'Inoperative - NCV, Broken -
Connector',
       'Not Latched - Door', 'Leak - Machine',
       'Open - Door, Broken - No Component Mentioned',
       'Broken - Plunger, Hit - Plunger', 'Leak - Sprayer',
       'Oil Leak - Sprayer', 'Stuck - Fuel Tank',
       'Loose - No Component Mentioned, Loose - Boom', 'Frozen -
Tank',
       'Oil Leak - Pipe',
```

```
       'Inoperative - Light, Inoperative - Light, Inoperative - Tank,
Inoperative - Light',
       'Error Code - No Component Mentioned', 'Blown - Air
Conditioner',
       'Blown - Hose', 'Not Mentioned - Sprayer', 'Incorrect - Panel',
       'No Cold Air - Air Conditioner',
       'Deflated - Air Bag, Inoperative - Washer', 'Inoperative -
Pump',
       'Not Mentioned - Boom, Not Mentioned - Boom, Protruded - Wing,
Inconsistent Folding - Boom',
       'Not Installed - Boom', 'Bent - Boom, Cracked - Boom',
       'Missed - Air Line, Leak - Plumbing System, Air Leak -
Connector',
       'Missed - Boom, Broken - Boom', 'Shut Down - NCV',
       'Derated - Engine, Faulty - No Component Mentioned',
       'Not Blowing - Cab', 'Rough - Engine, Dead - Engine',
       'Loose - Condenser', 'Broken - Light', 'Detached - Fuel Door',
       'Not Opening - Valve', 'Leak - O Ring, Not Mentioned - Hose',
       'Not Mentioned - Receiver',
       'Not Mentioned - Boom, Not Raised - Boom',
       'Stripped Thread - Boom', 'Fuel Leak - Seperator',
       'Fuel Pressure Loss - Machine', 'Missed - Pad',
       'Inoperative - Sensor', 'Not Updated - Hood', 'Leak - Bin',
       'Not Mentioned - Chassis', 'Calibration Issue - Drive System',
       'Communication Loss - Not Mentioned, Not Mentioned - Not
Mentioned',
       'Inoperative - Suspension', 'Inoperative - Tarp',
       'Broken - Harness, Failed - Sensor', 'Error Code - Engine',
       'Leak - Tank, Dirty - Tube',
       'Broken - Windshield, Not Mentioned - Window', 'Leak - Axle',
       'Leak - Bed, Loose - Bed', 'Leak - Fuel Tank',
       'Missed - Planetary Gear, Missed - Thrust', 'Dirty - Tube',
       'Error Code - Machine', 'Stuck Open - Wheel',
       'Not Installed - Mat', 'Not Mentioned - Boom',
       'Not Mentioned - No Component Mentioned, Not Mentioned - DEF
System',
       'Oil Leak - Hose', 'Not Mentioned - Transmission',
       'Oil Leak - Hydraulic Block', 'Not Latched - Steering',
       'Leak - Shaft, Excessive End Play - Shaft',
       'Not Installed - Oil Cooler', 'Cracked - Tip', 'Bent - Boom',
       'Not Folding - Boom', 'Not Mentioned - Tire',
       'Not Installed - Mudflap', 'Fraying - Strap',
       'Stuck - Auger, Inoperative - Auger', 'Leak - Fitting',
       'Leak - Seal, Dirty - No Component Mentioned',
       'Broken - Sensor, Dangling - Wire', 'Dirty - Cab, Dirty - Cab',
       'Not Sealed - Door', 'Leak - Steering', 'Leak - Fan',
       'Worn - Brake', 'Not Mentioned - Wire', 'Not Installed -
Manifold',
       'Not Mentioned - VSN', 'Faulty - Engine', 'Oil Leak - Fan',
```

```
       'Not Engaged - PTO', 'Not Mentioned - Rod',
       'Not Mentioned - Frame',
       'Faulty - Intake, Error Code - No Component Mentioned, Error
Code - No Component Mentioned',
       'Leak - No Component Mentioned, Missed - Sprayer, Missed - O
Ring',
       'Broken - Boom', 'Error Code - NCV, Not Spraying - Nozzle',
       'Air Leak - Air Bag',
       'Not Installed - Tool Box, Not Installed - Hose',
       'Intermittent - Conveyor',
       'Not Folding - Boom, Slow - Boom, Out of Range - Sensor',
       'Missed - Housing', 'Inoperative - Light', 'Leak - Cylinder',
       'Rotted - Auger', 'Not Mentioned - Fan, Not Mentioned - Bush',
       'Shattered - Door', 'Inoperative - XRT', 'Missed - Bumper',
       'Excessive Wear - Cylinder, Broken - Clip', 'Frozen - Boom',
       'Inconsistent Rate - Pump',
       'Error - Pump, Error - Boom, Inoperative - NCV',
       'Inoperative - Nozzle', 'Low Voltage - XRT', 'Inoperative -
NCV',
       'Erratic - XRT, Not Settling - Boom', 'Derated - Boom',
       'Shut Down - Viper', 'Not Installed - Recirculation System',
       'Not Installed - Tire', 'Faulty - Boom, Warning - Boom',
       'Incorrect - Plate', 'Leak - Hydraulic System', 'Cracked -
Boom',
       'Rubbed - Wheel, Rust - Wheel, Not Sealed - Hydraulic Tank',
       'Slow - Suspension', 'Rubbed - Steering', 'Leak - Filter',
       'Inoperative - Mirror', 'Inoperative - Strobe',
       'Leak - Hydro Case', 'Leak - Unit, Water Intrusion -
Windshield',
       'Not Mentioned - Wheel', 'Missed - Mirror, Broken - Mirror',
       'Not Mentioned - Breakaway', 'Broken - Tarp',
       'Not Functioning - Clutch, Inoperative - Bin',
       'Faulty - DEF System', 'Broken - Swing', 'Oil Leak - Hub',
       'Not Cooling - Air Conditioner', 'Lost - Battery',
       'Missed - Suspension', 'Broken - Armrest',
       'Not Detected - Machine, Not Holding - Wheel',
       'Leak - Hydraulic System, Leak - Motor', 'Inoperative -
System',
       'Disconnected - No Component Mentioned',
       'Derated - Machine, Error Code - Machine', 'Inoperative -
Spinner',
       'Not Sealed - Door, Air Leak - Door',
       'Not Mentioned - Auto Greaser', 'Not Mentioned - Valve',
       'Not Increasing - Fan, Oil Leak - Valve',
       'Not Hitting - No Component Mentioned, Hunting - No Component
Mentioned, Over Hitting - No Component Mentioned, Rate Issue - No
Component Mentioned',
       'Oil Leak - No Component Mentioned', 'Low - Coolant',
       'Not Mentioned - Driveshaft', 'Not Mentioned - Battery',
```

```
       'Leak - Boom, Damaged - Handrail',
       'Not Installed - Light, Not Installed - Light',
       'Out of Range - Sensor, Not Mentioned - XRT, Locked - Boom',
       'Dirty - Machine', 'Noise - Differential, Not Moved - Unit',
       'Leak - Air Bag', 'Leak - Turbo, Leak - Turbo', 'Leak - Box',
       'Leak - Wheel', 'Blowing Hot - Air Conditioner',
       'Not Sliding - Axle', 'Broken - Valve', 'Vibrated - XRT',
       'Error Code - ECM, Cut - Wire',
       'Not Spraying - No Component Mentioned', 'Not Raised - Mast',
       'Not Mentioned - Unit', 'GPS Issue - Antenna',
       'Leak - Cylinder, Leak - Cylinder',
       'Damaged - Drive Line, Cracked - Transfer Case, Cracked -
Engine, Not Mentioned - Drive Line, Cracked - Transmission',
       'Oil Leak - Hydraulic Pump',
       'Rough - Machine, Blocked - Port, Blocked - Screen',
       'Oil Leak - Pump',
       'Damaged - Intake, Damaged - Elbow, Damaged - Clamp',
       'Oil Leak - Cab', 'Low - Tank',
       'Damaged - Boom, Not Torqued - Fastener',
       'Leak - No Component Mentioned',
       'Blown - Air Conditioner, Not Mentioned - Steering',
       'Leak - Motor', 'Not Mentioned - Sensor',
       'Not Retaining - Battery', 'Creeped - Belt',
       'Inoperative - Switch',
       'Worn - Steering, Missed - No Component Mentioned',
       'Broken - Antenna, Stripped Thread - Antenna', 'Faulty -
Wiper',
       'Inoperative - Light, Inoperative - Light, Inoperative -
Flasher, Bent - Flasher, Broken - Fender, Inoperative - Beacon',
       'Not Speeding - Fan, Surging - Floor', 'Faulty - Air
Conditioner',
       'Leak - Pump', 'Not Updated - Drive Shaft', 'Damaged - Ladder',
       'Not Installed - Tarp',
       'Not Mentioned - Crank Case, Not Mentioned - Rack',
       'Not Mentioned - Block, Inoperative - Blinker, Not Mentioned -
Door, Not Mentioned - Steering, Not Mentioned - Filter',
       'Not Turning - Motor, Not Turning - Motor, Rate Issue -
Spreader, Calibration Issue - Spreader',
       'Not Controlling - Fan',
       'Pressure Issue - Pump, Not Turning On - Pump',
       'Not Installed - Slingshot', 'Not Installed - Camera',
       'Not Mentioned - Hood', 'Bent - Handle, Bent - Bracket',
       'Bent - Ladder', 'Not Increasing - Bin, Incorrect Reading -
Bin',
       'Not Installed - First Aid Kit, Not Installed - Extinguisher',
       'Not Seated - Machine, Not Mentioned - No Component Mentioned',
       'Loose - Chain, Bent - Gear Box', 'Not Blowing - Air
Conditioner',
       'Speed Issue - Valve', 'Not Installed - Boom, Broken - Nozzle',
```

```
       'Not Mentioned - Spray System', 'Not Mentioned - Filter',
       'Not Installed - Console', 'Fraying - Belt', 'Cracked - Hood',
       'Not Mentioned - Axle, Not Mentioned - Transfer Case, Not
Mentioned - Hub',
       'Not Gripped - Door, Broken - Door', 'Air Leak - Valve',
       'Worn - Intake', 'Moved - Steering',
       'Incorrect Reading - No Component Mentioned',
       'Incorrect Position - Console', 'Seized - Boom',
       'Black Out - Monitor', 'Ineffective - Bin',
       'Not Updated - No Component Mentioned', 'Not Mentioned -
Gearbox',
       'Leak - Hydraulic Pump, Leak - Motor',
       'Disconnected - Light, Out - Light, Not Seated - BIn',
       'Faulty - No Component Mentioned', 'Cracked - Valve', 'Leak -
PTO',
       'Faulty - Sensor, Faulty - DEF System',
       'Not Moved - Spinner, Not Moved - Spinner', 'Inoperative -
Radio',
       'Poor Quality - Shaft, Poor Quality - Drive Shaft',
       'Ruptured - Radiator', 'Leak - Auger',
       'Loose - No Component Mentioned, Loose - No Component
Mentioned',
       'Bent - Bracket, Missed - Bumper',
       'Broken - Spring, Broken - Spring', 'Not Mentioned - Door',
       'Not Spraying - NCV',
       'Error Code - DEF System, Derated - DEF System', 'Torn - Belt',
       'Not Shifted - Transmission', 'Broken - Cylinder',
       'Oil Leak - Cooler', 'Erratic - Breakaway', 'Leak - Pan',
       'Not Mentioned - Slack', 'Broken - Spring',
       'Worn - Air Conditioner', 'Not Mentioned - Shaft',
       'Worn - Steering', 'Stopped - Fan', 'Not Setting - Spinner',
       'Not Opening - Bin', 'Missed - Ladder',
       'Faulty - Wheel, Speed Restricted - Machine',
       'Blown - Engine, Dead - No Component Mentioned',
       'High Voltage - Lift', 'Not Engaged - Steering', 'Stuck -
Valve',
       'Not Turning - Starter', 'Oil Leak - Belt',
       'Burnt - Wire, Missed - Wire', 'Broken - Mirror, Noise -
Mirror',
       'Failed - Hydraulic System',
       'Not Engaged - Transfer Case, Not Mentioned - Four Wheel
Drive',
       'Derated - DEF System, Inoperative - DEF',
       'Smashed - Hood, Smashed - Grill', 'Leak - Spinner',
       'Not Installed - Tarp, Fraying - Tarp, Damaged - Tarp',
       'Stuck - Axle', 'Not Reading - DEF System',
       'Not Installed - No Component Mentioned, Not Installed - Drain,
Not Installed - First Aid Kit, Not Installed - Extinguisher',
       'Broken - Leaf, Shifted - Spring, Twisted - Spring',
```

```
       'Not Mentioned - Hose, Not Mentioned - Boom',
       'Rate Issue - Machine', 'Worn - Fan', 'Leak - Fuel Filter',
       'Leak - Differential', 'Not Mentioned - Drive Shaft',
       'Stuck - Sensor', 'Not Spraying - Nozzle', 'Hit - Shield',
       'Not Folding - Boom, Protruded - Boom', 'Broken - Breakaway',
       'Error Code - Viper', 'Not Mounted - Camera',
       'Not Seated - Boom, Not Achieving - Boom',
       'Not Starting - Machine', 'Incorrect Reading - Fuel Tank',
       'Control Issue - Venturi', 'Oil Leak - Engine', 'Leak - Belt',
       'Not Turning - Auger', 'Incorrect Reading - Fan',
       'Melted - Switch', 'Alarmed - No Component Mentioned',
       'Loose - Gasket', 'Leak - Auxiliary Transmission',
       'Hydraulic Issue - Sensor',
       'Rate Issue - Machine, Chain Issue - Machine, Faulty - Bin',
       'Error Code - Unit', 'No Start - Machine'], dtype=object)
```

```
merged_df['Failure Condition - Failure Component'] =
merged_df['Failure Condition - Failure Component'].str.strip()

# Strip and simplify failure labels to the first condition
merged_df['Primary Failure Component'] = merged_df['Failure Condition
- Failure Component'].str.split(',').str[0].str.strip()

top_primary_failures = merged_df['Primary Failure
Component'].value_counts().nlargest(8).index
filtered_df = merged_df[merged_df['Primary Failure
Component'].isin(top_primary_failures)]

merged_df['Primary Failure
Component'].value_counts(dropna=False).head(10)
```

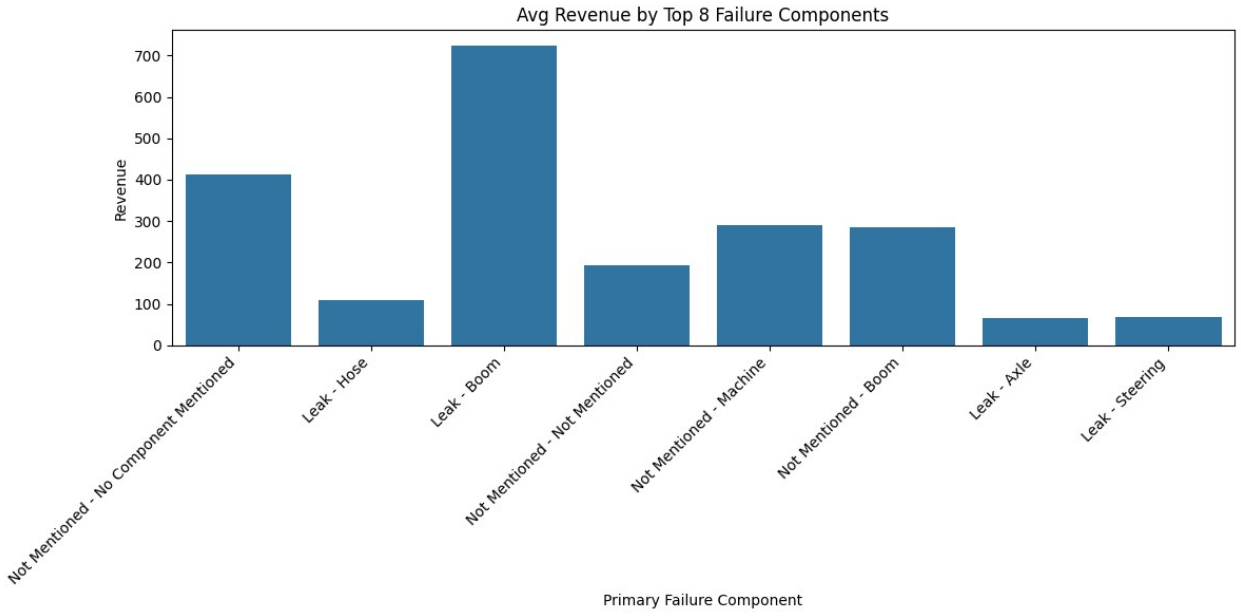
```
Primary Failure Component
Not Mentioned - Not Mentioned             72
Leak - Hose                                7
Leak - Steering                            7
Leak - Boom                                6
Not Mentioned - Boom                       6
Not Mentioned - Machine                    6
Leak - Axle                                6
Not Mentioned - No Component Mentioned     6
Not Folding - Boom                         4
Inoperative - Light                        4
Name: count, dtype: int64
```

```
merged_df[['Primary Failure Component', 'Cost']].dropna().shape
```

```
(0, 2)
```

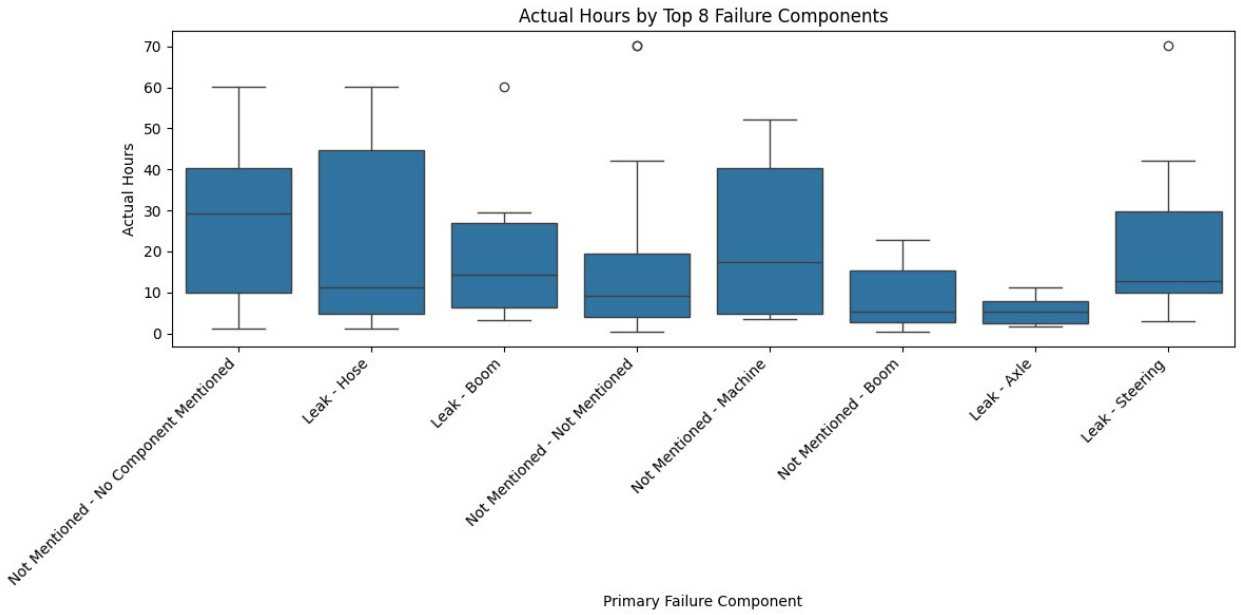
Plot: Avg Revenue by Failure Component (Top 8)

```
plt.figure(figsize=(12,6))
sns.barplot(data=filtered_df,
            x='Primary Failure Component',
            y='Revenue',
            estimator='mean',
            errorbar=None)
plt.title("Avg Revenue by Top 8 Failure Components")
plt.xticks(rotation=45, ha='right')
plt.tight_layout()
plt.show()
```



Plot: Actual Hours by Failure Component (Boxplot)

```
plt.figure(figsize=(12,6))
sns.boxplot(data=filtered_df,
            x='Primary Failure Component',
            y='Actual Hours')
plt.title("Actual Hours by Top 8 Failure Components")
plt.xticks(rotation=45, ha='right')
plt.tight_layout()
plt.show()
```




```
merged_df.to_excel("merged_task2_output.xlsx", index=False)
```

Executive Summary: Task 2 & Task 3

Task 2: Data Preparation and Integration

This task focused on cleaning, formatting, and joining two datasets for further analysis.

Steps Performed:

- **Loading Data**: Two Excel sheets were loaded into `df1` and `df2`.
- **Join Strategy**: Both datasets contained a common column `"Primary Key"`, which was used to perform a **left join**, keeping all records from `df1`.
- **Data Cleaning**:
  - Checked for and noted any `null` values.
  - Dropped duplicate records if present.
  - Stripped whitespace from column names to standardize naming.
  - Converted columns like `Cost` and `Segment Total` to float data types.
  - Filled missing values:
    - `Coverage` → replaced with `"Unknown"`.
    - `Cause`, `Correction` in `df1` → filled with `"Not Mentioned"`.

Task 3.1 – Trend Analysis

This analysis involved identifying temporal patterns and data relationships through visual exploration.

Steps Performed:

- **Date Handling**: Ensured `Order Date` was converted to proper datetime format.
- **Monthly Trends**: Grouped data by month and counted records to assess trends.
- **Visualizations**:
  - Line plot to show monthly record count.
  - Heatmap showing **correlation between numeric variables**, with interpretation of:
    - Pearson Correlation Coefficient scale (from -1 to 1).
    - Use of Coolwarm color map to distinguish strength and direction of relationships.

Task 3.2 – Root Cause Identification

This sub-task explored how component failures contribute to revenue and resource consumption.

Steps Performed:

- **Average Revenue by Failure Component**:
  - Bar chart showing the top 8 failure components contributing the most revenue loss.
- **Actual Hours by Failure Component**:
  - Boxplot showing variation in actual hours spent across different failure components.

Summary of Key Actions

- Standardized and merged two datasets via primary key.
- Cleaned and preprocessed text, numeric, and date fields.
- Used visual tools (line plot, heatmap, bar chart, boxplot) to explore:
  - Monthly trends
  - Variable relationships
  - Revenue and time distribution by failure component