US Supply Chain Risk Analysis

Presented by:

Saran Chandrasekharan Unnithan

Content

Dataset Description

Data Explanation

Data Summarization

Hypothesis and Test Results

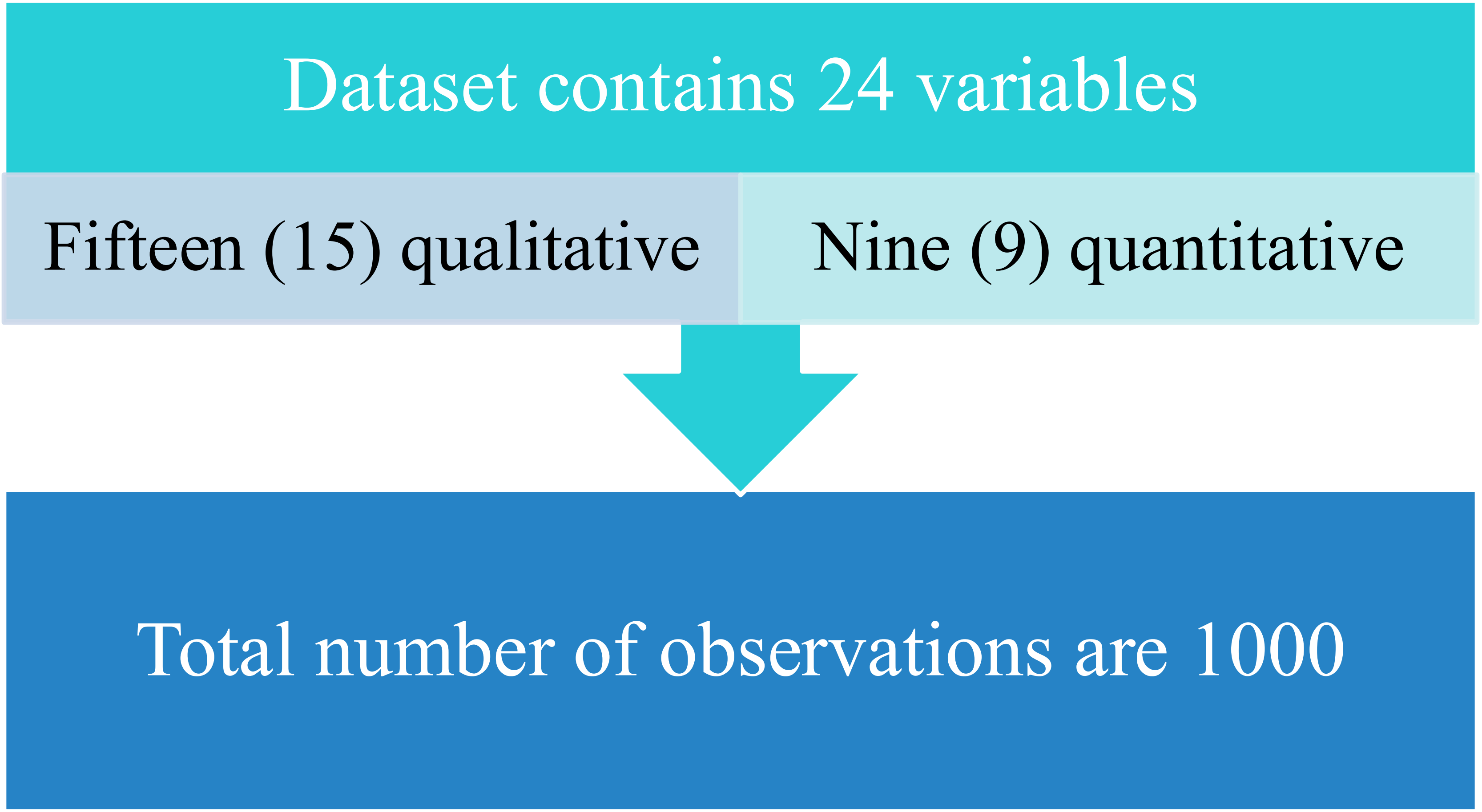
Conclusion

Supplement Slides

Dataset Description

- Sourced from Kaggle.com
- CSV File
- US Supply Chain Risk Analysis (2023)
- Objective

Dataset Explanation



Dataset Summarization- Numerical

MEASURES	QUANTITY ORDERED	ORDER VALUE (USD)	COMMUNICATION COST (MB)	ENERGY CONSUMPTION (JOULES)	HISTORICAL DISRUPTION COUNT	SUPPLIER RELIABILITY SCORE	AVAILABLE HISTORICAL RECORDS	FEDERATED ROUND	PARAMETER CHANGE MAGNITUDE
Mean	504.288	25290.073	2.768	270.855	9.841	0.745	510.317	5.679	0.251
Median	499	25596.005	2.76	268.37	10	0.74	521.5	6	0.255
Min - Max	10- 989	500.58- 49444.6	0.5- 5	50.49- 449.35	0- 19	0.5- 1	10- 988	1- 10	0.001- 0.488
Standard deviation	295.545	14380.329	1.275	130.218	5.656	0.1430	281.700	2.864	0.139
Skewness	0.026	-0.041	-0.043	0.0415	-0.070	0.012	-0.020	-0.031	0.002
Kurtosis	1.763	1.788	1.906	1.8	1.845	1.799	1.791	1.785	1.825
IQR	512.5	24682.592	2.082	220.237	10	0.25	489.25	5	0.237
Q1	248.00	12891.838	1.74	163.59	5.00	0.62	262.25	3.00	0.01
Q3	761.50	37607.41	3.83	384.01	15.00	0.87	752.00	8.00	0.37
Lower Bound	-522.25	-24132.051	-1.39	-167.03	-10	0.25	-472.38	-4.50	-0.224
Upper Bound	1531.75	74598.319	6.96	714.62	30	1.25	1486.23	15.50	0.726

Data Summarization – Categorical

Product Category

- Textile-206, Electronics-210, Machinery- 204, Food- 190, Pharma- 190

Shipping Mode

- Air- 263, Rail- 255, Road- 256, Sea- 226

Delay Days

- 0- 486, 1- 133, 2- 105, 3- 99, 5- 70, 7- 39, 10- 68

Disruption Type

- None-486, Shortage- 135, Weather- 133, Customs- 124, Strike- 122

Disruption Severity

- None- 486, Low- 248, Medium- 152, High- 114

Dominant Buyer Flag

- 0's- 714, 1's- 286

Data Sharing Consent

- 0's- 195, 1's- 805

Supplier Risk Flag

- 0's- 486, 1's- 514

Data Cleaning

- No missing values
- No Duplicate data
- No special characters
- No extreme values
- No data cleaning

Hypothesis Tests and Variables

Test	Variables
Z-Test	Product Category(Electronics) and Order Value
Two-Sample t-Test	Supply Risk Flag and Supplier Reliability Score
One-Way ANOVA	Energy Consumption by Shipping Mode (Air, Road & Rail)
Two-Way ANOVA	Shipping Mode, Disruption Severity & Energy Consumption.
Chi^2 Test	Disruption Severity and Shipping Mode
Linear Regression	Order Value and Quantity Ordered
Multiple Linear Regression	Order Value, Quantity Ordered, Historical Disruption Count, Supplier Reliability Score and Communication Cost

Z- Test

Hypothesis

The average order value for the Electronics product category is **significantly higher** than the overall average order value.

Assumption Testing

Normality: Shapiro–Wilk p-values $6.666*10^{-07}$ → data **not normally distributed**.

Post Logarithmic Normality Test: Shapiro–Wilk p-values $5.338*10^{-15}$ → data **not normally distributed**.

Z test results

Z= 1.8144, p-value= 0.03481

α- 0.01- not significant

α- 0.05- significant

α- 0.10- significant

Conclusion

There is **moderate evidence** that Electronics product orders are higher than average, but there is a **lack of strong statistical evidence (1%)**

T- Test

Hypothesis	Suppliers with supplier risk flag 0 have ***significantly higher*** reliability scores.
Assumption Testing	**Normality:** Shapiro–Wilk p-values 0- $3.524*10^{-10}$ and 1- $1.982*10^{-12}$ → data **not normally distributed**. **Post Logarithmic Normality Test:** Shapiro–Wilk p-values for 0-$9.14*10^{-11}$ and for 1-$5.882*10^{-13}$ → data **not normally distributed**.
Homogeneity of Variance Test	p- value 0.1061 >0.05 (homogeneity of variance met)
T- test results	**T= -1.4955, p-value= 0.9324>0.05**
Conclusion	Suppliers with supplier risk flag 0 ***do not have significantly higher*** reliability scores.

One Way ANOVA

Hypothesis	Air, Road, and Rail shipping modes **have similar** energy consumption.
Assumption Testing	**Normality:** Shapiro–Wilk p-values < 0.05 for all modes → data **not normally distributed**. **Post Logarithmic Normality Test:** Shapiro–Wilk p-values <0.05→ data **not normally distributed**.
Homogeneity of Variance Test	Levene's Test p = 0.7233>0.05 → **equal variances assumed**.
ANOVA Results	p = 0.582 → **no significant difference** in mean energy consumption across modes.
Post-hoc (Tukey HSD)	All adjusted p-values > 0.05 → **no pairwise differences** between Air, Road, and Rail.
Conclusion	Shipping mode has **no statistically significant impact** on energy consumption.

Two Way ANOVA

Hypothesis

H_1: The energy consumption significantly impacts air, rail, and road shipping modes, and high and low disruption severities

H_2: The energy consumption significantly impacts high and low disruption severities

$H3$: There is an interaction effect between shipping mode and disruption severity.

Assumption Testing

Normality: Shapiro–Wilk p-values p-value<0.05 → data **not normally distributed**.

Post Logarithmic Normality Test: Shapiro–Wilk p-value < 0.05 → data **not normally distributed**.

Homogeneity of Variance Test

Levene's Test p = 0.232 > 0.05 → **equal variances assumed**.

Two-way ANOVA Results

p = 0.968>0.05 for shipping mode,

p = 0.559> 0.05 for Disruption Severity.

p = 0.297>0.05 for their interaction → **no significant difference** shipping modes, disruption severities or their interaction

Post-hoc (Tukey HSD)

All adjusted p-values > 0.05 → **no pairwise differences** between Air, Road, and Rail.

Conclusion

No statistical significance of energy consumption on shipping modes, disruption severity and on their interaction

Chi- Square test

Hypothesis

Disruption severity depends on shipping mode.

Chi^{2}Results

χ^2= 4.63859, df=9 and p- value = 0.864

α- 0.01- not significant

α- 0.05- not significant

α- 0.10- not significant

Conclusion

No statistical significance of energy consumption on shipping modes, disruption severity and on their interaction

Simple Linear Regression

Hypothesis

Quantity Ordered does have a linear effect on Order Value USD

Model Findings

p-value = 0.492, above all conventional significance thresholds (1%, 5%, 10%).

R^2 = 0.00047, indicating the model explains almost none of the variation.

Overall regression is not statistically significant.

Interpretation

The relationship between the two variables is **extremely weak**.

The predictor offers **no meaningful explanatory power** for the outcome.

Conclusion

No evidence of a significant linear effect in this dataset.

Multiple Linear Regression

Hypothesis

Quantity Ordered, Historical Disruption Count, Supplier Reliability Score, Communication Cost **has significant linear effect** on Order Value

Model Findings

p-value = **Quantity Ordered** is 0.429, **Historical Disruption Count** is 0.784, **Supplier Reliability Score** is 0.385, **Communication Cost** is 0.171 , above all conventional significance thresholds (1%, 5%, 10%).

Adjuster R^2 = 0.00075, indicating the model explains almost none of the variation.

Interpretation

Overall regression is not statistically significant

The predictor offers no meaningful explanatory power for the outcome.

VIF (Variance Inflation Factor)

The variance inflation factor (VIF) values for all predictors are approximately 1→ **no presence of multicollinearity**

Conclusion

No evidence of a significant linear effect in this dataset.

Hypothesis Tests and Results

Test	Variables	p-value	Results
Z-Test	Product Category(Electronics) and Order Value	0.0348	Significant at 5% & 10%, not at 1%
Two-Sample t-Test	Supply Risk Flag and Supplier Reliability Score	0.9324	Homogeneity of variance is supported(0.1061). Not Significant at 1%, 5%, 10%.
One-Way ANOVA	Energy Consumption by Shipping Mode (Air, Road & Rail)	0.582	Homogeneity of variance is supported(0.7233). Not Significant at 1%, 5%, 10%.
Two-Way ANOVA	Shipping Mode, Disruption Severity & Energy Consumption.	Shipping Mode: 0.968	Homogeneity of variance is supported(0.232). Not Significant at 1%, 5%, 10%.
		Disruption Severity:0.559	Not Significant at 1%, 5%, 10%.
		Combined: 0.297	Not Significant at 1%, 5%, 10%.
Two-way Chi^2 Test	Disruption Severity and Shipping Mode	0.8646	Not Significant at 1%, 5%, 10%.
Linear Regression	Order Value and Quantity Ordered	0.492	$R^2 = 0.00047$(0.047%) Not Significant at 1%, 5%, 10%.
Multiple Linear Regression	Order Value, Quantity Ordered, Historical Disruption Count, Supplier Reliability Score and Communication Cost	Quantity ordered: 0.429	$R^2 = 0.0032$(0.32%), The VIF values for all predictors are approximately 1, indicating no multicollinearity. Not Significant at 1%, 5%, 10%.
		Hist Dis: 0.784	
		Supp Rel:0.385	
		Comm:0.171	

Conclusion

Shipping mode shows **no significant effect** on **energy** consumption.

Disruption severity is **independent** of **shipping** mode.

Regression models show **no meaningful predictors** of order value.

Dataset signals are **weak, limiting statistical significance** across tests.

Recommendation: Increase data variability and integrate product effects to improve model significance

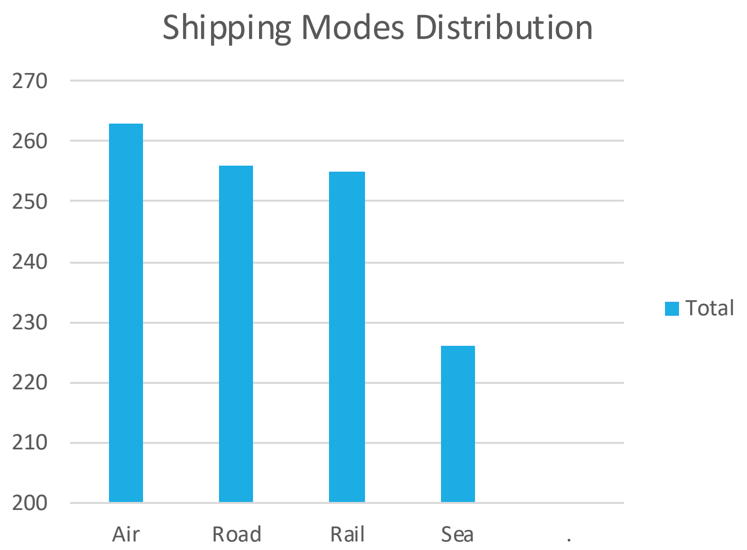
Thank you!

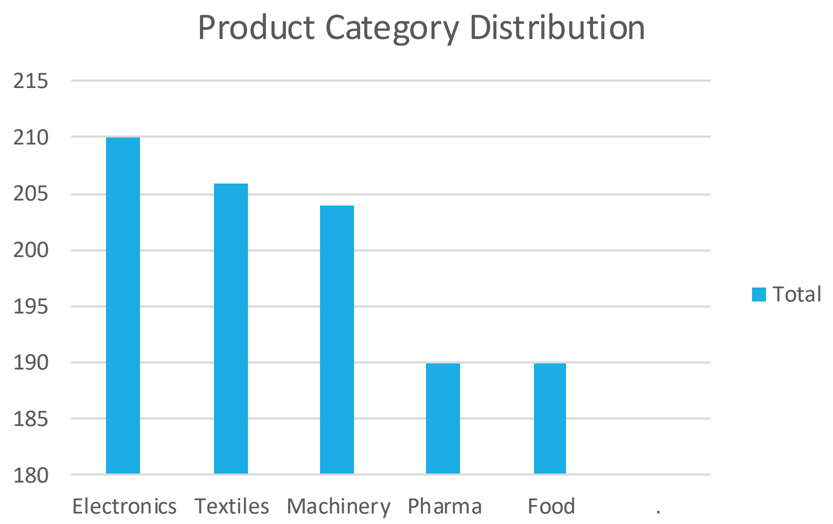
Supplemental Slides

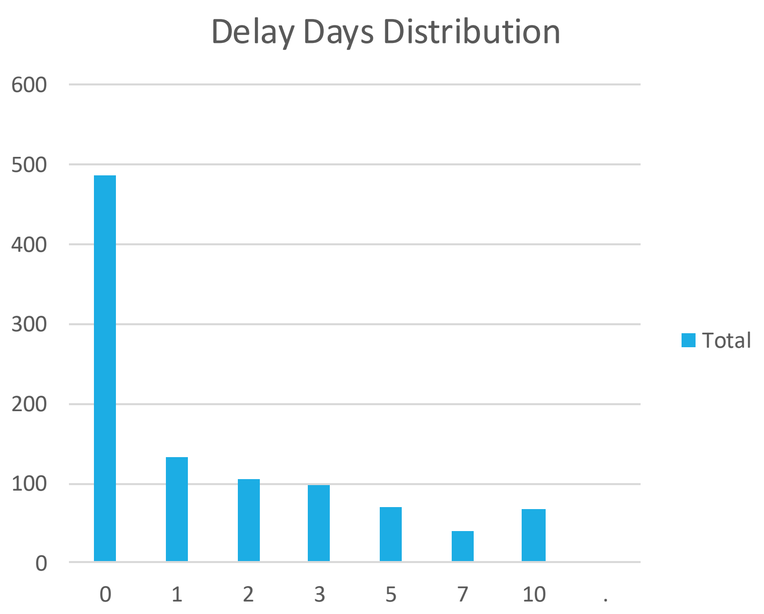
Dataset

Order_ID	Buyer_ID	Supplier_ID	Product_Category	Quantity_Ordered	Order_Date	Dispatch_Date	Delivery_Date	Shipping_Mode	Order_Value_USD	Delay_Days	Disruption_Type	Disruption_Severity	Historical_Disruption_Count	Supplier_Reliability_Score	Organization_ID	Dominant_Buyer_Flag	Available_Historical_Records	Data_Sharing_Consent	Federated_Round	Parameter_Change_Magnitude	Communication_Cost_MB	Energy_Consumption_Joules	Supply_Risk_Flag
O1000	B33	S23	Textiles	469	10/24/23	10/27/23	10/28/23	Rail	36273.99	0	None	None	11	0.73	Org13	0	127	1	10	0.0225	2.27	236.06	0
O1001	B1	S20	Machinery	365	7/7/23	7/8/23	7/9/23	Road	34780.36	0	None	None	1	0.88	Org8	1	909	1	7	0.0412	3.1	257.8	0
O1002	B2	S10	Food	333	12/28/23	12/29/23	1/7/24	Rail	7154.54	7	Shortage	Low	19	0.95	Org4	0	262	1	8	0.1183	2.82	165.38	1
O1003	B6	S10	Machinery	142	1/14/23	1/17/23	1/20/23	Rail	15320.08	0	None	None	17	0.92	Org18	1	807	1	9	0.2611	3.59	377.47	0
O1004	B5	S4	Machinery	897	1/12/23	1/14/23	1/16/23	Road	18256.42	0	None	None	5	0.6	Org13	0	789	1	7	0.2775	3.42	83.1	0
O1005	B9	S11	Pharma	814	3/28/23	3/30/23	4/8/23	Rail	40319.64	0	None	None	9	0.72	Org14	0	812	1	10	0.0268	2.41	150.24	0
O1006	B34	S30	Food	849	1/12/23	1/14/23	1/23/23	Road	14298.61	0	None	None	19	0.6	Org16	0	854	1	10	0.0935	2.41	386.85	0
O1007	B7	S30	Food	136	1/25/23	1/27/23	2/5/23	Air	10929.9	0	None	None	14	0.63	Org12	1	189	1	8	0.2549	1.13	246.71	0
O1008	B16	S24	Food	115	10/28/23	10/29/23	11/7/23	Sea	47893.68	0	None	None	4	0.75	Org18	1	424	1	9	0.301	2.81	367.38	0
O1009	B26	S13	Machinery	613	8/31/23	9/1/23	9/10/23	Sea	938.73	1	Shortage	High	19	0.53	Org17	0	879	1	6	0.0827	3.14	305.37	1
O1010	B30	S25	Textiles	808	5/2/23	5/5/23	5/8/23	Rail	49892.13	0	None	None	2	0.61	Org6	0	244	1	10	0.2054	2.64	477.82	0
O1011	B30	S14	Food	387	5/27/23	5/31/23	6/5/23	Rail	34002.65	0	None	None	4	0.97	Org13	0	864	1	2	0.4532	1.97	319.01	0
O1012	B42	S3	Machinery	914	2/24/23	2/28/23	3/7/23	Rail	41509.23	0	None	None	7	0.73	Org15	1	717	1	2	0.4995	3.75	56.69	0
O1013	B49	S8	Pharma	60	8/26/23	8/28/23	9/5/23	Sea	15083.66	1	Weather	Low	7	0.72	Org17	0	829	0	4	0.4937	0.84	208.79	1
O1014	B3	S5	Textiles	38	12/22/23	12/23/23	1/1/24	Air	1208.61	7	Customs	Low	7	0.51	Org13	1	89	1	9	0.442	4.61	114.38	1
O1015	B34	S9	Pharma	471	6/22/23	6/24/23	6/27/23	Air	37024.54	0	None	None	11	0.62	Org2	1	656	1	7	0.2034	0.59	466.09	0
O1016	B7	S27	Textiles	305	8/9/23	8/10/23	8/13/23	Sea	41790.16	1	Customs	Low	7	0.93	Org17	0	31	1	2	0.3464	4.83	96.81	1
O1017	B17	S16	Machinery	306	4/29/23	4/30/23	5/7/23	Air	37153.57	0	None	None	10	0.58	Org14	1	486	1	8	0.2141	2.58	224.47	0
O1018	B22	S8	Electronics	735	11/5/23	11/7/23	11/14/23	Sea	7571.09	2	Customs	Low	13	0.71	Org3	0	852	1	7	0.3278	2.36	151.74	1
O1019	B36	S18	Textiles	788	7/2/23	7/4/23	7/12/23	Air	37794.68	5	Customs	Medium	7	0.51	Org12	0	449	1	9	0.1616	1.65	431.02	1
O1020	B37	S24	Pharma	416	10/15/23	10/17/23	11/1/23	Road	38561.68	10	Strike	High	9	0.74	Org4	0	746	1	8	0.307	0.55	188.85	1
O1021	B31	S9	Machinery	778	11/3/23	11/6/23	11/12/23	Road	33097.76	0	None	None	6	0.51	Org11	0	579	0	3	0.2666	1.74	496.35	0
O1022	B50	S21	Textiles	823	10/27/23	10/30/23	11/5/23	Sea	38422.74	2	Weather	Low	16	0.91	Org14	1	102	1	1	0.022	0.99	67.38	1
O1023	B41	S12	Electronics	622	3/23/23	3/24/23	4/3/23	Air	42373.11	7	Customs	Low	3	0.91	Org11	0	201	1	2	0.3974	3.47	107.72	1
O1024	B23	S21	Electronics	414	7/30/23	8/3/23	8/12/23	Rail	30873.4	1	Strike	Low	12	0.89	Org16	0	112	0	2	0.0158	1.94	151.09	1
O1025	B6	S7	Food	355	5/30/23	6/2/23	6/6/23	Air	4885.93	3	Strike	Low	3	0.64	Org8	0	405	0	4	0.1536	3.95	415	1
O1026	B37	S23	Food	631	2/6/23	2/10/23	2/14/23	Road	24637.58	3	Strike	Low	4	0.92	Org9	0	16	0	6	0.0129	1.43	445.33	1
O1027	B23	S10	Pharma	429	12/18/23	12/20/23	12/27/23	Road	4343.64	0	None	None	3	0.77	Org18	0	649	1	4	0.1943	4.55	82.15	0
O1028	B14	S7	Textiles	191	11/8/23	11/12/23	11/18/23	Air	20673.38	2	Strike	Low	17	0.69	Org9	0	576	1	9	0.0409	2.85	64.61	1
O1029	B39	S16	Textiles	992	12/27/23	12/30/23	1/8/24	Air	20651.77	7	Shortage	High	5	0.81	Org16	0	240	0	4	0.4576	1.02	385.05	1
O1030	B34	S3	Food	834	2/28/23	3/2/23	3/13/23	Rail	3767.49	7	Strike	Low	1	0.51	Org11	1	883	1	6	0.2137	2.3	110.01	1
O1031	B35	S16	Textiles	10	5/11/23	5/14/23	5/18/23	Rail	17766.62	1	Customs	Low	10	0.57	Org3	1	583	0	6	0.2352	1.43	109.85	1
O1032	B2	S11	Textiles	840	7/21/23	7/25/23	7/29/23	Air	5994.41	2	Customs	High	13	0.87	Org20	0	249	1	10	0.1814	1.4	60.06	1
O1033	B4	S2	Electronics	959	8/31/23	9/4/23	9/9/23	Air	40507.64	0	None	None	4	0.62	Org2	1	99	1	10	0.3175	2.78	72.41	0
O1034	B41	S20	Food	576	2/4/23	2/8/23	2/10/23	Air	47410.56	0	None	None	5	0.8	Org5	0	241	1	1	0.3944	3.52	168.65	0
O1035	B32	S16	Machinery	380	4/3/23	4/7/23	4/12/23	Air	4079.67	0	None	None	9	0.69	Org19	0	466	1	8	0.3276	4.6	79.25	0

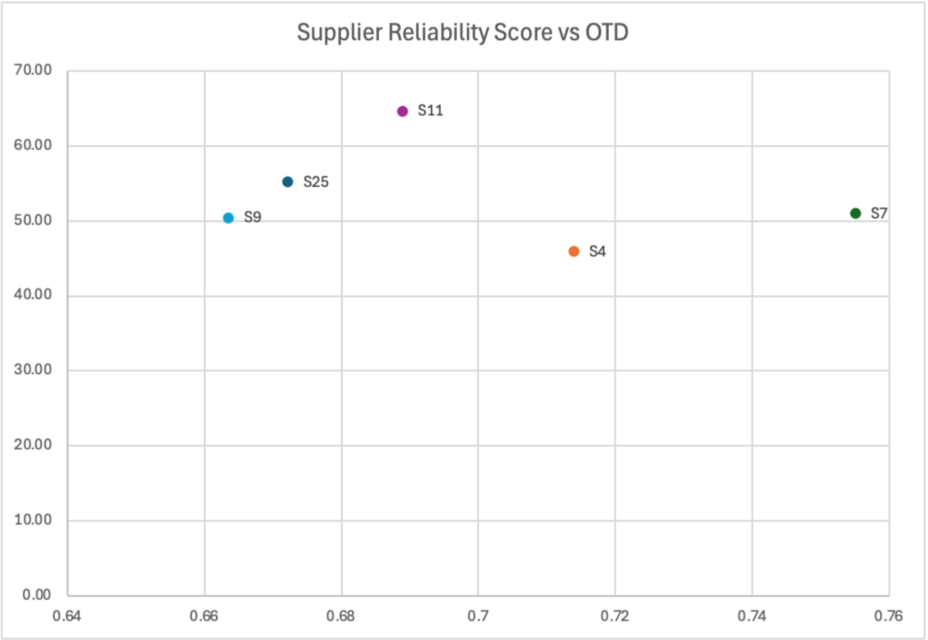
Important Visualizations







Delay Days Distribution
600
500
400
300
200
100
0
0
1
2
3
5
7
10
Total



Supplier Reliability Score vs OTD
70.00
60.00
50.00
40.00
30.00
20.00
10.00
0.00
0.64
0.66
0.68
0.7
0.72
0.74
0.76
S11
S25
S9
S7
S4

Python Code for Descriptive Statistics

```
from google.colab import files
import pandas as pd

# Upload file
uploaded = files.upload()
df = pd.read_csv("Group7_USSupplyChainRiskAnalysis.csv")
display(df.head())

# Python code descriptive statistics for each numerical variable

import numpy as np

# List of numerical columns to analyze
numerical_cols = [
    'Quantity_Ordered',
    'Order_Value_USD',
    'Communication_Cost_MB',
    'Energy_Consumption_Joules',
    'Historical_Disruption_Count',
    'Supplier_Reliability_Score',
    'Available_Historical_Records',
    'Federated_Round',
    'Parameter_Change_Magnitude'
]

# Calculating the descriptive statistics for each numerical column
for col in numerical_cols:
    print(f"--- Statistics for {col} ---")

    # Central Tendencies
    mean_val = df[col].mean()
    median_val = df[col].median()
    mode_val = df[col].mode().tolist() # Get all modes
```

```
print(f"Mean: {mean_val:.2f}")
print(f"Median: {median_val:.2f}")
print(f"Mode: {mode_val}")

# Measures of Dispersion
std_dev = df[col].std()
skewness = df[col].skew()
kurtosis = df[col].kurtosis()

print(f"Standard Deviation: {std_dev:.2f}")
print(f"Skewness: {skewness:.2f}")
print(f"Kurtosis: {kurtosis:.2f}")

# Quartiles, IQR, Bounds, Min, Max
quartiles = df[col].quantile([0.25, 0.5, 0.75])
Q1 = quartiles[0.25]
Q3 = quartiles[0.75]
IQR = Q3 - Q1

lower_bound = Q1 - 1.5 * IQR
upper_bound = Q3 + 1.5 * IQR

min_val = df[col].min()
max_val = df[col].max()

print(f"Q1 (25th percentile): {Q1:.2f}")
print(f"Q3 (75th percentile): {Q3:.2f}")
print(f"IQR (Interquartile Range): {IQR:.2f}")
print(f"Lower Bound (for outliers): {lower_bound:.2f}")
print(f"Upper Bound (for outliers): {upper_bound:.2f}")
print(f"Minimum: {min_val:.2f}")
print(f"Maximum: {max_val:.2f}")
print("-" * 30)
```

```
--- Statistics for Quantity_Ordered ---
Mean: 504.29
Median: 499.00
Mode: [327]
Standard Deviation: 295.55
Skewness: 0.03
Kurtosis: -1.24
Q1 (25th percentile): 248.00
Q3 (75th percentile): 760.50
IQR (Interquartile Range): 512.50
Lower Bound (for outliers): -520.75
Upper Bound (for outliers): 1529.25
Minimum: 10.00
Maximum: 999.00
```

```
--- Statistics for Communication_Cost_MB ---
Mean: 2.77
Median: 2.76
Mode: [1.87]
Standard Deviation: 1.28
Skewness: -0.04
Kurtosis: -1.09
Q1 (25th percentile): 1.74
Q3 (75th percentile): 3.82
IQR (Interquartile Range): 2.08
Lower Bound (for outliers): -1.38
Upper Bound (for outliers): 6.95
Minimum: 0.50
Maximum: 5.00
```

```
--- Statistics for Order_Value_USD ---
Mean: 25290.07
Median: 25596.01
Mode: [500.58, 506.67, 567.0, 572.95, 598.61, 689.09, 715.98, 763.02, 888.16, 911.84, 914.0, 938.73
Standard Deviation: 14380.33
Skewness: -0.04
Kurtosis: -1.21
Q1 (25th percentile): 12891.84
Q3 (75th percentile): 37574.43
IQR (Interquartile Range): 24682.59
Lower Bound (for outliers): -24132.05
Upper Bound (for outliers): 74598.32
Minimum: 500.58
Maximum: 49945.18
```

```
--- Statistics for Energy_Consumption_Joules ---
Mean: 270.86
Median: 268.37
Mode: [58.35, 102.38, 116.31, 120.16, 128.61, 150.24, 165.19, 183.52, 208.24, 270.66, 317.55,
Standard Deviation: 130.22
Skewness: 0.04
Kurtosis: -1.20
Q1 (25th percentile): 163.60
Q3 (75th percentile): 383.83
IQR (Interquartile Range): 220.24
Lower Bound (for outliers): -166.76
Upper Bound (for outliers): 714.19
Minimum: 50.49
Maximum: 499.84
```

```
--- Statistics for Historical_Disruption_Count ---
Mean: 9.84
Median: 10.00
Mode: [12]
Standard Deviation: 5.66
Skewness: -0.07
Kurtosis: -1.16
Q1 (25th percentile): 5.00
Q3 (75th percentile): 15.00
IQR (Interquartile Range): 10.00
Lower Bound (for outliers): -10.00
Upper Bound (for outliers): 30.00
Minimum: 0.00
Maximum: 19.00
```

```
--- Statistics for Supplier_Reliability_Score ---
Mean: 0.75
Median: 0.74
Mode: [0.66]
Standard Deviation: 0.14
Skewness: 0.01
Kurtosis: -1.20
Q1 (25th percentile): 0.62
Q3 (75th percentile): 0.87
IQR (Interquartile Range): 0.25
Lower Bound (for outliers): 0.24
Upper Bound (for outliers): 1.25
Minimum: 0.50
Maximum: 1.00
```

```
--- Statistics for Available_Historical_Records ---
Mean: 510.32
Median: 521.50
Mode: [405, 529, 715, 725]
Standard Deviation: 281.70
Skewness: -0.02
Kurtosis: -1.21
Q1 (25th percentile): 262.75
Q3 (75th percentile): 752.00
IQR (Interquartile Range): 489.25
Lower Bound (for outliers): -471.12
Upper Bound (for outliers): 1485.88
Minimum: 10.00
Maximum: 998.00
```

```
--- Statistics for Federated_Round ---
Mean: 5.68
Median: 6.00
Mode: [10]
Standard Deviation: 2.86
Skewness: -0.03
Kurtosis: -1.22
Q1 (25th percentile): 3.00
Q3 (75th percentile): 8.00
IQR (Interquartile Range): 5.00
Lower Bound (for outliers): -4.50
Upper Bound (for outliers): 15.50
Minimum: 1.00
Maximum: 10.00
```

```
--- Statistics for Parameter_Change_Magnitude ---
Mean: 0.25
Median: 0.26
Mode: [0.1659, 0.2516, 0.3468, 0.4937]
Standard Deviation: 0.14
Skewness: 0.00
Kurtosis: -1.18
Q1 (25th percentile): 0.13
Q3 (75th percentile): 0.37
IQR (Interquartile Range): 0.24
Lower Bound (for outliers): -0.22
Upper Bound (for outliers): 0.73
Minimum: 0.01
Maximum: 0.50
```

Buyer Positioning

Buyer ID	Count of Buyer_ID
B12	31
B40	30
B29	29
B34	27
B47	26
B4	25
B23	25
B22	25
B6	24
B30	24

Dominant_Buyer_Flag	1		
Row Labels	**Count of Supplier_ID**	**Supplier Base**	**Dominance Ratio**
B16	12	22	54.54
B12	9	31	29.03
B34	9	27	33.33
B45	9	23	39.13
B39	8	20	40
B41	8	18	44.44
B1	8	20	40
B29	8	29	27.58

Supplier Positioning

Supplier ID	Count of Supplier_ID
S24	48
S25	45
S16	44
S4	42
S7	42
S26	42
S10	39
S23	39
S1	37
S9	35
S28	35
S29	35

	Order_Value_USD	On-time delivery
Textiles	594469.34	
S9	214878.72	20.58958722
S25	197516.74	17.65801917
S19	182073.88	23.13331105
Electronics	560920.66	
S30	208365.7	31.92305351
S7	179086.66	17.0403556
S4	173468.3	14.16355512
Pharma	515953.04	
S25	181685.3	16.24268662
S4	170500.27	13.92121772
S20	163767.47	23.56676946
Machinery	453410.39	
S16	153758.65	14.9675412
S22	152987.54	18.98512118
S7	146664.2	13.95531147
Food	387982.07	
S16	141449.52	13.76931651
S1	127781.01	14.35922596
S24	118751.54	9.529394068

Supplier Optimization

Row Labels	Average of Parameter_Change_Magnitude	Sum of Federated_Round	Average of Supplier_Reliability_Score
S24	0.266	267	0.72
S25	0.244	264	0.725
S26	0.256	257	0.717
S16	0.272	256	0.752
S7	0.274	247	0.772

Best Supplier

Delay_Days	0			
Supplier ID	**Average of Parameter_Change_Magnitude**	**Sum of Federated_Round**	**Sum of Order_Value_USD**	**Average of Supplier_Reliability_Score**
S25	0.227	136	617127.63	0.725
S4	0.271	103	562706.25	0.714
S7	0.302	149	535689.97	0.755
S9	0.261	98	526246.36	0.664
S11	0.2	113	499096.01	0.689

Best Shipping Mode

Delay_Days	0			
Row Labels	**Average of Communication_Cost_MB**	**Sum of Order_Value_USD**	**Average of Energy_Consumption_Joules**	**Count of Shipping_Mode**
Air	2.632	3619116.23	250.638	138
Rail	2.743	2881662.86	263.042	118
Road	2.890	3209890.64	265.644	125
Sea	2.985	2620264.4	286.429	105

Order Quantity

- Orders are generally centered near **500 units**.
- The most common order size (**327 units**) is smaller than the average.
- Dispersion of 295.5 defines a **high variability in order size**.
- Distribution is almost **perfectly symmetric** (no strong tilt).
- Spread is wide, with orders ranging from **10 to 999 units**.
- The flatter curve (low kurtosis) means **fewer extreme peaks**.
- Overall, the data with aggregate 504,288 reflects **diverse ordering behaviors** rather than clustering around one level.

Key takeaway: The Quantity_Ordered variable is centered around ~500 units, **fairly symmetrical**, and **highly dispersed**. The combination of a flat distribution and wide variability suggests **multiple procurement behaviours** (some buyers consistently place medium orders, while others swing between very small and very large quantities).

Order Value in USD

•Average order values hover around **$25K**.

•Median is slightly higher than the mean, hinting at a **slight left skew**.

•No single order value dominates.

•**Standard Deviation:** 14,380 defines a large spread in order values.

•Large spread: orders range from **$500 to nearly $50K**.

•The flatter distribution indicates **fewer extreme outliers**.

•Overall, spending patterns are **broad but balanced**, without major anomalies.

Key Takeaway : Order_Value_USD is a **well-balanced, symmetric variable** with **high variability** and **no extreme outliers dominating the distribution**.

- •The pattern suggests a mix of **high-value bulk orders and smaller routine purchases**, providing insights into buyer segmentation by order size.

Communication Cost in MB

•Average cost is **~2.77 MB**, with values typically centered near **2.76 MB**.

•The most common cost (**1.87 MB**) is slightly below the average.

•Variation is moderate (**1.28 MB**) compared to the small mean, showing noticeable spread.

•The range (**0–4.5 MB**) indicates both very light and heavy communication instances.

•Distribution is nearly symmetric (skew = **–0.04**), with no strong tilt.

•Kurtosis (**1.91**) suggests a flatter-than-normal distribution—fewer sharp peaks.

Key Takeaway: Communication costs are fairly balanced, moderately variable, and without heavy extremes

Energy Consumption in Joules

•Centered near **271 J**, with median **268 J**.

•Multiple common values (modes like **424, 485, 389 J, etc.**) reflect diverse consumption patterns.

•Spread is wide: from **0 to 449 J**.

•High variability (**130 J**) shows strong fluctuation in usage.

•Distribution is symmetric (skew = **0.04**), not tilted strongly.

•Kurtosis (**1.8**) reflects a flat, dispersed pattern.

Key Takeaway: Energy consumption is widely spread, highly variable, and reflects multiple consumption regimes rather than a single cluster.

Historical Disruption Count

•Centered near **10 disruptions**, both mean (**9.84**) and median (**10**) align closely.

•Mode is **12**, slightly above the center.

•Range is broad (**0–19** disruptions).

•Standard deviation (**5.66**) highlights strong variability across history.

•Symmetry holds (skew = **–0.07**).

•Kurtosis (**1.85**) shows a flatter spread—less frequent extremes.

Key Takeaway: Historical disruptions average around 10 but vary significantly, showing both stable and unstable supply episodes.

Supplier Reliability Score

- Average reliability is **0.745**, close to median (**0.74**).
- Mode is **0.66**, slightly lower.
- Range is **0.5–1.0**, covering a wide reliability span.
- Moderate variability (**0.14**).
- Symmetric distribution (skew = **0.01**).
- Low kurtosis (**1.80**) indicates no extreme concentration.

Key Takeaway: Supplier reliability is fairly consistent around ~0.75, but spread suggests both less reliable and near-perfect suppliers exist.

Available Historical Records

•Very large average records (**510**), with median **522**.
•Common record levels cluster around **405–725**, showing distinct groups.
•Spread is extreme (**range = 988**).
•Huge variability (**282**) relative to the mean.
•Distribution is symmetric (skew = **–0.02**).
•Flat distribution (**kurtosis = 1.79**) indicates no sharp clustering.

Key Takeaway: Historical records availability is highly dispersed, reflecting significant differences in data completeness across sources.

Federated Round

- Mode is **10**, higher than central tendency, showing frequent later rounds.
- Spread is moderate (**range = 9**, **SD = 2.86**).
- Mean round is **5.68**, very close to median (**6**).
- Almost symmetric (skew = **–0.03**).
- Low kurtosis (**1.79**) indicates flatter spread across rounds.

Key Takeaway: Federated learning rounds cluster around 5–6, but with noticeable dispersion and frequent late rounds.

Parameter Change Magnitude

- Centered at **0.25**, with median **0.255**.
- Modes cluster at multiple small values (**0.166–0.494**).
- Range (**0–0.49**) indicates variation from minimal to substantial change.
- Dispersion is moderate (**SD = 0.14**).
- Symmetry (skew = **0.002**) suggests no tilt.
- Flat kurtosis (**1.83**) confirms broad spread.

Key Takeaway: Parameter changes are usually small (~0.25) but vary widely, with no extreme concentration.

Mode for Quantitative Variables

Variables	Mode
Quantity Ordered	327
Order Value (USD)	#N/A
Communication Cost (MB)	1.87
Energy Consumption (Joules)	423.97, 484.7, 428.89, 388.61, 486.94, 486.16, 317.55, 270.66, 208.24, 183.52, 150.24, 128.61, 58.35, 165.19,102.38, 120.16, 116.31
Historical Disruption Count	12
Supplier Reliability Score	0.66
Available Historical Records	405, 529, 715, 725
Federated Round	10
Parameter Change Magnitude	0.1659, 0.2516, 0.3468, 0.4937

Mode for Qualitative Variables

Variables	Mode
Order ID	
Supplier ID	S24
Buyer ID	B12
Order Date	
Dispatch Date	
Delivery Date	
Product Category	Electronics
Shipping Mode	Air
Delay Days	0
Disruption Types	None
Disruption Severity	None
Organization ID	Org11
Dominant Buyer Flag	0
Data Sharing Consent	1
Supply Risk Flag	1

Test Results

Z-Test

```
> Electro=filter(US_Supply_Chain,Product_Category=="Electronics")
> shapiro.test(Electro$Order_Value_USD)

	Shapiro-Wilk normality test

data:  Electro$Order_Value_USD
W = 0.94774, p-value = 6.666e-07

> Electro$log10_Order_Value <- log(Electro$Order_Value_USD, base = 10)
> shapiro.test(Electro$log10_Order_Value)

	Shapiro-Wilk normality test

data:  Electro$log10_Order_Value
W = 0.81657, p-value = 5.338e-15
```

```
> library(BSDA)
> z.test(x = Electro$Order_Value_USD,mu = 25290,sigma.x = 14380,alternative = "greater")

	One-sample z-Test

data:  Electro$Order_Value_USD
z = 1.8144, p-value = 0.03481
alternative hypothesis: true mean is greater than 25290
95 percent confidence interval:
 25458.21       NA
sample estimates:
mean of x 
 27090.42 
```

T-Test

```
        Shapiro-Wilk normality test

data:  Low_risk$Supplier_Reliability_Score
W = 0.96024, p-value = 3.524e-10
```

```
        Shapiro-Wilk normality test

data:  High_risk$Supplier_Reliability_Score
W = 0.9482, p-value = 1.982e-12
```

```
        Shapiro-Wilk normality test

data:  Low_risk$log10_Supplier_Reliability_Score
W = 0.9566, p-value = 9.14e-11
```

```
        Shapiro-Wilk normality test

data:  High_risk$log10_Supplier_Reliability_Score
W = 0.94442, p-value = 5.882e-13
```

```
Levene's Test for Homogeneity of Variance (center = median)
       Df F value Pr(>F)
group   1   2.616 0.1061
      998
```

T-Test Ctd.

```
        Two Sample t-test

data:  Supplier_Reliability_Score by as.factor(Supply_Risk_Flag)
t = -1.4955, df = 998, p-value = 0.9324
alternative hypothesis: true difference in means between group 0 and group 1 is greater than 0
95 percent confidence interval:
 -0.02842133         Inf
sample estimates:
mean in group 0 mean in group 1 
      0.7388066       0.7523346 
```

One Way ANOVA

```
> Air=filter(US_Supply_Chain,Shipping_Mode=="Air")
> Rail=filter(US_Supply_Chain,Shipping_Mode=="Rail")
> Road=filter(US_Supply_Chain,Shipping_Mode=="Road")
> shapiro.test(Air$Energy_Consumption_Joules)

	Shapiro-Wilk normality test

data:  Air$Energy_Consumption_Joules
W = 0.95331, p-value = 1.821e-07

> shapiro.test(Rail$Energy_Consumption_Joules)

	Shapiro-Wilk normality test

data:  Rail$Energy_Consumption_Joules
W = 0.96076, p-value = 2.004e-06

> shapiro.test(Road$Energy_Consumption_Joules)

	Shapiro-Wilk normality test

data:  Road$Energy_Consumption_Joules
W = 0.95069, p-value = 1.289e-07
```

```
> Air$log10_Energy_Consumption_Joules<- log(Air$Energy_Consumption_Joules, base = 10)
> Rail$log10_Energy_Consumption_Joules<- log(Rail$Energy_Consumption_Joules, base = 10)
> Road$log10_Energy_Consumption_Joules<- log(Road$Energy_Consumption_Joules, base = 10)
> shapiro.test(Air$log10_Energy_Consumption_Joules)

	Shapiro-Wilk normality test

data:  Air$log10_Energy_Consumption_Joules
W = 0.93238, p-value = 1.334e-09

> shapiro.test(Rail$log10_Energy_Consumption_Joules)

	Shapiro-Wilk normality test

data:  Rail$log10_Energy_Consumption_Joules
W = 0.92027, p-value = 1.903e-10

> shapiro.test(Road$log10_Energy_Consumption_Joules)

	Shapiro-Wilk normality test

data:  Road$log10_Energy_Consumption_Joules
W = 0.92966, p-value = 1.121e-09
```

One Way ANOVA Ctd.

```
> leveneTest(Energy_Consumption_Joules ~ as.factor(Shipping_Mode), data = USSupplyChain)
Levene's Test for Homogeneity of Variance (center = median)
       Df F value Pr(>F)
group   2  0.3241 0.7233
      771

> ShippingMode_Anova<-aov(Energy_Consumption_Joules ~ as.factor(Shipping_Mode), data = USSupplyC
ain)
> summary(ShippingMode_Anova)
                          Df   Sum Sq Mean Sq F value Pr(>F)
as.factor(Shipping_Mode)   2    18063    9031   0.543  0.582
Residuals                771 12835105   16647
```

```
> TukeyHSD(ShippingMode_Anova)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Energy_Consumption_Joules ~ as.factor(Shipping_Mode), data = USSupplyChain)

$`as.factor(Shipping_Mode)`
               diff       lwr      upr     p adj
Rail-Air  10.767377 -15.86026 37.39501 0.6090516
Road-Air   9.521406 -17.07981 36.12262 0.6779264
Road-Rail -1.245972 -28.05222 25.56028 0.9934536
```

Two Way ANOVA

```
> Air=filter(df,Shipping_Mode=="Air")
> Rail=filter(df,Shipping_Mode=="Rail")
> Road=filter(df,Shipping_Mode=="Road")
> shapiro.test(Air$Energy_Consumption_Joules)

	Shapiro-Wilk normality test

data:  Air$Energy_Consumption_Joules
W = 0.94615, p-value = 0.001319

> shapiro.test(Rail$Energy_Consumption_Joules)

	Shapiro-Wilk normality test

data:  Rail$Energy_Consumption_Joules
W = 0.9511, p-value = 0.001589

> shapiro.test(Road$Energy_Consumption_Joules)

	Shapiro-Wilk normality test

data:  Road$Energy_Consumption_Joules
W = 0.94124, p-value = 0.0003953
```

```
> Air$log10_Energy_Consumption_Joules<- log(Air$Energy_Consumption_Joules, base = 10)
> Rail$log10_Energy_Consumption_Joules<- log(Rail$Energy_Consumption_Joules, base = 10)
> Road$log10_Energy_Consumption_Joules<- log(Road$Energy_Consumption_Joules, base = 10)
> shapiro.test(Air$log10_Energy_Consumption_Joules)

	Shapiro-Wilk normality test

data:  Air$log10_Energy_Consumption_Joules
W = 0.88937, p-value = 2.205e-06

> shapiro.test(Rail$log10_Energy_Consumption_Joules)

	Shapiro-Wilk normality test

data:  Rail$log10_Energy_Consumption_Joules
W = 0.90877, p-value = 7.521e-06

> shapiro.test(Road$log10_Energy_Consumption_Joules)

	Shapiro-Wilk normality test

data:  Road$log10_Energy_Consumption_Joules
W = 0.91057, p-value = 9.182e-06
```

Two Way ANOVA Ctd.

```
> leveneTest(Energy_Consumption_Joules ~ Shipping_Mode * Disruption_Severity, data = df)
Levene's Test for Homogeneity of Variance (center = median)
       Df F value Pr(>F)
group   5  1.3801  0.232
      266
```

```
> model <- aov(Energy_Consumption_Joules ~ Shipping_Mode * Disruption_Severity, data = df)
> summary(model)
                                   Df  Sum Sq Mean Sq F value Pr(>F)
Shipping_Mode                       2    1092     546   0.032  0.968
Disruption_Severity                 1    5783    5783   0.341  0.559
Shipping_Mode:Disruption_Severity   2   41361   20681   1.221  0.297
Residuals                         266 4505069   16936
```

Chi-Square Test

```
> tbl <- table(df$Disruption_Severity, df$Shipping_Mode)
> print(tbl)

         Air Rail Road Sea
  High    26   29   32  27
  Low     60   64   61  63
  Medium  39   44   38  31
  None   138  118  125 105
> chi_result <- chisq.test(tbl)
> chi_result

	Pearson's Chi-squared test

data:  tbl
X-squared = 4.6386, df = 9, p-value = 0.8646

> chi_result$expected

             Air    Rail    Road     Sea
  High    29.982  29.07  29.184  25.764
  Low     65.224  63.24  63.488  56.048
  Medium  39.976  38.76  38.912  34.352
  None   127.818 123.93 124.416 109.836
> chi_contrib

              High          Low       Medium
  Air  0.5288614502 0.4184069668 0.0238286972
  Rail 0.0001685587 0.0091334598 0.7084004128
  Road 0.2717192982 0.0975010081 0.0213750000
  Sea  0.0592957615 0.8623020268 0.3270815091

              None
  Air  0.8110995634
  Rail 0.2837480836
  Road 0.0027412551
  Sea  0.2129255982
> 
```

Linear Regression

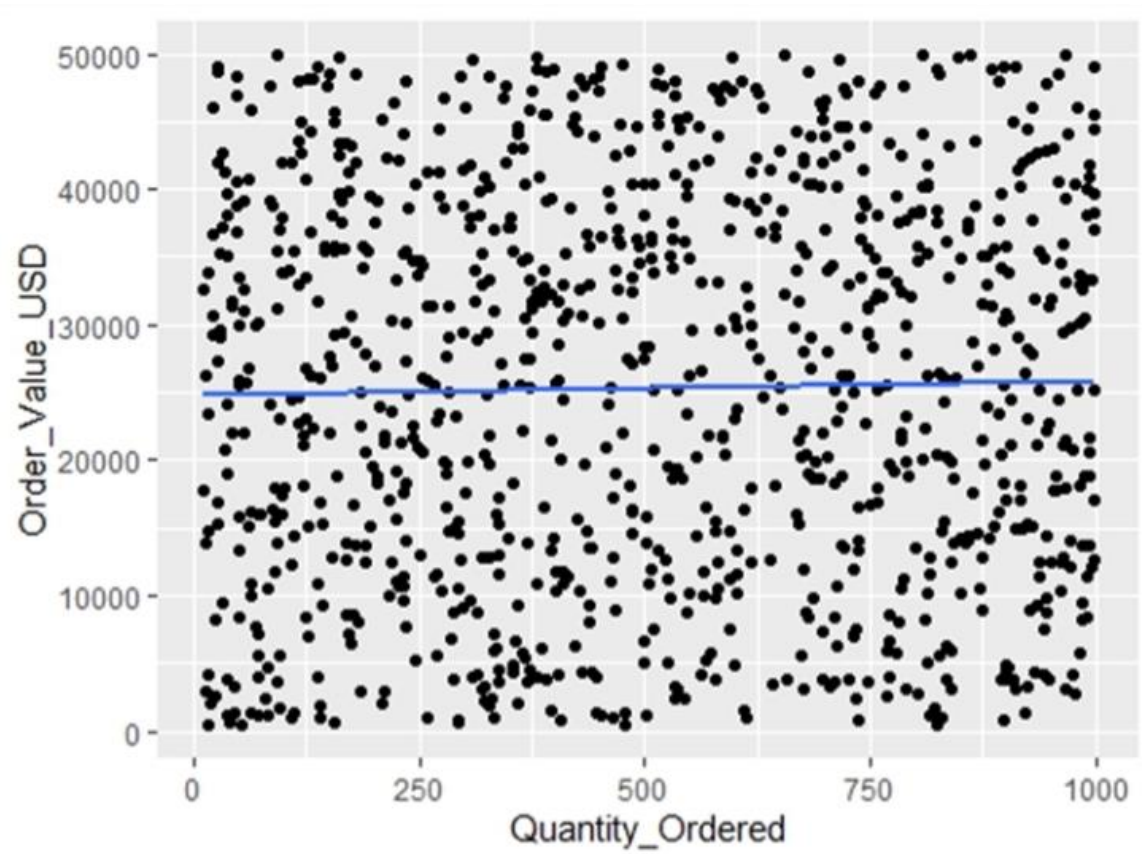
```
Call:
lm(formula = Order_Value_USD ~ Quantity_Ordered, data = US_Supply_Chain)

Residuals:
     Min       1Q   Median       3Q      Max 
-25055.4 -12204.9    445.8  12207.7  25008.4 

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)    
(Intercept)      24756.429    899.941  27.509   <2e-16 ***
Quantity_Ordered     1.058      1.540   0.687    0.492    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 14380 on 998 degrees of freedom
Multiple R-squared:  0.000473,	Adjusted R-squared:  -0.0005285 
F-statistic: 0.4723 on 1 and 998 DF,  p-value: 0.4921
```

Linear Regression Plot



Multi-Linear Regression

```
Call:
lm(formula = Order_Value_USD ~ Quantity_Ordered + Historical_Disruption_Count +
    Supplier_Reliability_Score + Communication_Cost_MB, data = data)

Residuals:
     Min       1Q   Median       3Q      Max
-26391.2 -12221.7    308.9  12216.5  25364.0

Coefficients:
                             Estimate Std. Error t value Pr(>|t|)
(Intercept)                 28313.534   2802.830  10.102   <2e-16 ***
Quantity_Ordered                1.222      1.546   0.791    0.429
Historical_Disruption_Count   -22.058     80.504  -0.274    0.784
Supplier_Reliability_Score  -2767.892   3182.590  -0.870    0.385
Communication_Cost_MB        -490.660    358.139  -1.370    0.171
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 14390 on 995 degrees of freedom
Multiple R-squared:  0.003247,	Adjusted R-squared:  -0.0007599
F-statistic: 0.8104 on 4 and 995 DF,  p-value: 0.5186
```

```
vif(model)
           Quantity_Ordered Historical_Disruption_Count  Supplier_Reliability_Score       Communication_Cost_MB
                   1.007262                    1.001034                    1.000726                    1.006873
```

Linear Regression_Dummy Variable

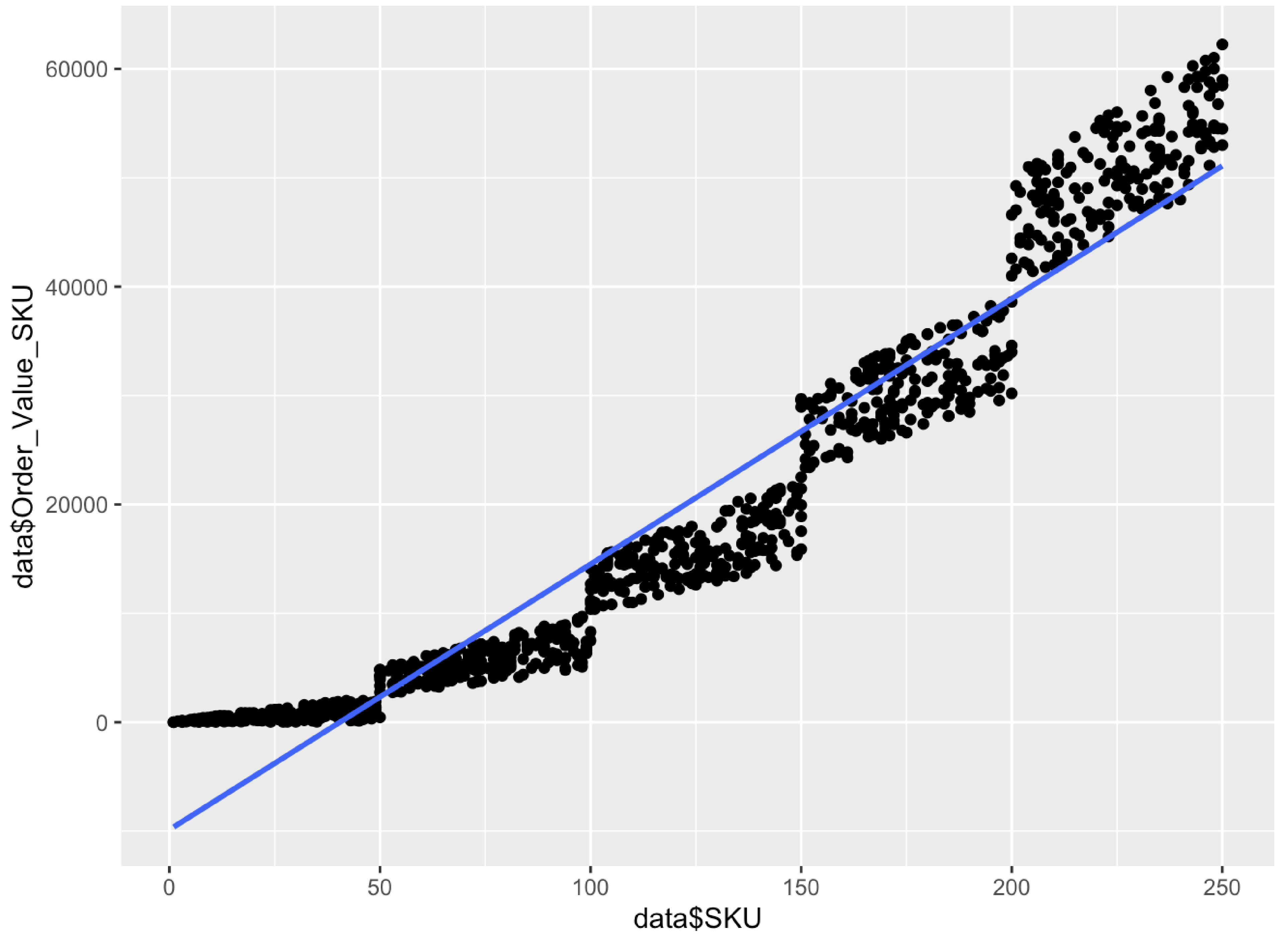
```
> model_lm <- lm(data$Order_Value_SKU ~ data$SKU, data = data)
> summary(model_lm)

Call:
lm(formula = data$Order_Value_SKU ~ data$SKU, data = data)

Residuals:
     Min       1Q   Median       3Q      Max 
-11119.1  -4047.8   -504.6   3582.8  11328.6 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -9861.601    316.621  -31.15   <2e-16 ***
data$SKU      243.810      2.232  109.23   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5071 on 998 degrees of freedom
Multiple R-squared:  0.9228,	Adjusted R-squared:  0.9227 
F-statistic: 1.193e+04 on 1 and 998 DF,  p-value: < 2.2e-16
```


60000
40000
20000
0
0
50
100
150
200
250
data$SKU
data$Order_Value_SKU

R Code for Z Test

Z_test.R

```
library(moments)
?mean
#import package for data wrangling
library(tidyverse)
getwd()
Electro=filter(US_Supply_Chain,Product_Category=="Electronics")
shapiro.test(Electro$Order_Value_USD)
Electro$log10_Order_Value <- log(Electro$Order_Value_USD, base = 10)
shapiro.test(Electro$log10_Order_Value)
install.packages("BSDA")
library(BSDA)
z.test(x = Electro$Order_Value_USD,mu = 25290,sigma.x = 14380,alternative = "greater")
```

R Code for T Test

T-Test.R

```
Low_risk=filter(US_Supply_Chain,Supply_Risk_Flag=="0")
High_risk=filter(US_Supply_Chain,Supply_Risk_Flag=="1")
shapiro.test(Low_risk$Supplier_Reliability_Score)
shapiro.test(High_risk$Supplier_Reliability_Score)
Low_risk$log10_Supplier_Reliability_Score <- log(Low_risk$Supplier_Reliability_Score, base = 10)
High_risk$log10_Supplier_Reliability_Score <- log(High_risk$Supplier_Reliability_Score, base = 10)
shapiro.test(Low_risk$log10_Supplier_Reliability_Score)
shapiro.test(High_risk$log10_Supplier_Reliability_Score)
leveneTest(Supplier_Reliability_Score ~ as.factor(Supply_Risk_Flag), data = US_Supply_Chain)
t.test(Supplier_Reliability_Score ~ as.factor(Supply_Risk_Flag),data = US_Supply_Chain,alternative="greater", var.equal = TRUE)
```

R Code for One Way ANOVA

One-Way_ANOVA.R

```r
#One-way ANOVA
Air=filter(US_Supply_Chain,Shipping_Mode=="Air")
Rail=filter(US_Supply_Chain,Shipping_Mode=="Rail")
Road=filter(US_Supply_Chain,Shipping_Mode=="Road")
Sea=filter(US_Supply_Chain,Shipping_Mode=="Sea")
shapiro.test(Air$Energy_Consumption_Joules)
shapiro.test(Rail$Energy_Consumption_Joules)
shapiro.test(Road$Energy_Consumption_Joules)
shapiro.test(Sea$Energy_Consumption_Joules)
Air$log10_Energy_Consumption_Joules<- log(Air$Energy_Consumption_Joules, base = 10)
Rail$log10_Energy_Consumption_Joules<- log(Rail$Energy_Consumption_Joules, base = 10)
Road$log10_Energy_Consumption_Joules<- log(Road$Energy_Consumption_Joules, base = 10)
Sea$log10_Energy_Consumption_Joules<- log(Sea$Energy_Consumption_Joules, base = 10)
shapiro.test(Air$log10_Energy_Consumption_Joules)
shapiro.test(Rail$log10_Energy_Consumption_Joules)
shapiro.test(Road$log10_Energy_Consumption_Joules)
shapiro.test(Sea$log10_Energy_Consumption_Joules)
leveneTest(Energy_Consumption_Joules ~ as.factor(Shipping_Mode), data = US_Supply_Chain)
ShippingMode_Anova<-aov(Energy_Consumption_Joules ~ as.factor(Shipping_Mode), data = US_Supply_Chain)
summary(ShippingMode_Anova)
TukeyHSD(ShippingMode_Anova)

```

R Code for Two Way ANOVA

Two-way ANOVA.R

```
library(tidyverse)
library(dplyr)
library(car)
library(WRS2)
library(onewaytests)
library(lmPerm)
getwd()
df <- read_csv("US Supply Chain.csv") %>%
  filter(Shipping_Mode %in% c("Air", "Road", "Rail"),Disruption_Severity %in% c("Low","High")) %>%
  mutate(Shipping_Mode = as.factor(Shipping_Mode),Disruption_Severity = as.factor(Disruption_Severity))
Air=filter(df,Shipping_Mode=="Air")
Rail=filter(df,Shipping_Mode=="Rail")
Road=filter(df,Shipping_Mode=="Road")
shapiro.test(Air$Energy_Consumption_Joules)
shapiro.test(Rail$Energy_Consumption_Joules)
shapiro.test(Road$Energy_Consumption_Joules)
Air$log10_Energy_Consumption_Joules<- log(Air$Energy_Consumption_Joules, base = 10)
Rail$log10_Energy_Consumption_Joules<- log(Rail$Energy_Consumption_Joules, base = 10)
Road$log10_Energy_Consumption_Joules<- log(Road$Energy_Consumption_Joules, base = 10)
shapiro.test(Air$log10_Energy_Consumption_Joules)
shapiro.test(Rail$log10_Energy_Consumption_Joules)
shapiro.test(Road$log10_Energy_Consumption_Joules)

leveneTest(Energy_Consumption_Joules ~ Shipping_Mode * Disruption_Severity, data = df)
model <- aov(Energy_Consumption_Joules ~ Shipping_Mode * Disruption_Severity, data = df)
summary(model)
t2way(Energy_Consumption_Joules ~ Shipping_Mode * Disruption_Severity, data = df)
tukey_results <- TukeyHSD(model)
tukey_results
```

R Code for Chi Square Test

```
getwd()
df <- read.csv("USSupplyChainRiskAnalysis.csv")
tbl <- table(df$Disruption_Severity, df$Shipping_Mode)
print(tbl)
chi_result <- chisq.test(tbl)
chi_result
chi_result$expected
observed <- chi_result$observed
expected <- chi_result$expected
chi_contrib <- (observed - expected)^2 / expected
chi_contrib
```

R Code for Linear Regression

Linear_Reg.R

```r
#linear regression
data <- read.csv("US Supply Chain Risk Analysis.csv")

model <- lm(Order_Value_USD ~ Quantity_Ordered, data = data)

summary(model)

library(ggplot2)
ggplot(data = data, aes(x=Quantity_Ordered, y=Order_Value_USD)) +
  geom_point() +
  geom_smooth(method = "lm", se=FALSE)
```

R Code for Multi Linear Regression

Multi-linear_Reg.R

```
# Multiple Linear Regression
model <- lm(Order_Value_USD ~ Quantity_Ordered +
              Historical_Disruption_Count +
              Supplier_Reliability_Score +
              Communication_Cost_MB,
            data = data)
# Residual normality
res <- residuals(model)

# QQ plot
qqnorm(res)
qqline(res, col = "red")

# Shapiro-Wilk test
shapiro.test(res)

model2 <- lm(log(Order_Value_USD) ~ Quantity_Ordered +
               Historical_Disruption_Count +
               Supplier_Reliability_Score +
               Communication_Cost_MB,
             data = data)
res2 <- residuals(model2)
shapiro.test(res2)


summary(model)
vif(model)
```