

SMART LOGISTICS ANALYTICS

Phase 1: Trade Logistics Data Analysis
– Problem Understanding & Preprocessing

Our Team



Priyansh Johri

PRN:24070126142



Shlok Panchal

PRN:24070126132



Nilraj Sinh Rana

PRN:214070126148

PROBLEM STATEMENT

- The logistics industry struggles to balance efficiency, cost, sustainability, and risk management.
- Shipments often face delays due to distance, carrier performance, or checkpoints.
- Rising transport, customs, and insurance costs affect profitability.
- Increasing fuel consumption and CO₂ emissions raise sustainability concerns.
- Shipments are at risk of damage from vibration, handling, temperature, and humidity factors.

OBJECTIVES

- Delay Prediction → Identify patterns in trade & transport modes that lead to shipment delays.
- Cost Optimization → Analyze CIF values & partner trade flows to reduce logistics expenses.
- Sustainability Goals – Evaluate the impact of trade on fuel consumption and emissions, recommending greener logistics practices.
- Risk Assessment → Detect anomalies in reported vs estimated data for fraud or error detection.

Market Overview

The global cargo logistics market is worth around USD 4–5.5 trillion (2024) and is expected to reach USD 6–8 trillion by 2030, growing at ~4–6% annually. Growth is driven by e-commerce, global trade, digitalization, and demand for fast delivery. Air cargo is expanding fastest (~7% yearly), while maritime cargo grows moderately (~3–5%). Key challenges include geopolitical risks, fuel costs, and sustainability pressures. Asia-Pacific leads the market due to strong trade and manufacturing hubs.



4.3%-5.6%
Market Growth Rate



USD 5.95 trillion
by 2030

Market Size



4.8% year-over-year
Growth for 2025

DATA COLLECTION

- 1** Source → Public trade dataset (import/export flows with countries, commodities, transport).
- 2** Granularity → Partner-country level with product category & trade flow.
- 3** Variables → Trade values (CIF, Primary), Quantities (qty, net weight), Flags (estimated, reported), Transport Mode.
- 4** Purpose → Provides raw material for analyzing international trade dynamics and logistics.

Raw Dataset

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Shipment_ID	Origin	Destination	Cargo_Type	Weight_kg	Volume_m3	Departure_Date	Arrival_Date	Carrier	Cost_USD	Status	Priority	Packaging_Type	Customer_ID
2	S1000	Port Christopherpc	Stephensmouth	Machinery	2611.87	42.86	30-06-2025	11-07-2025	RapidTrans	3723.47	In Transit	Low	Pallet	b67e9135-1bc7-4
3	S1001	Richardborough	Gomezton	Furniture	2248.51	63.46	08-07-2025	16-07-2025	FastTrack	3383.11	Delayed	Medium	Container	b15ddc3b-4ec6-4
4	S1002	West Jeffrey	Josephton	Pharmaceuticals	1675.53	14.03	27-07-2025	30-07-2025	CargoIndia	3197.83	Delivered	Medium	Box	b8f15326-dbce-4
5	S1003	Port Jenniferhaven	Michaelshire	Machinery		85.98	17-08-2025	22-08-2025	BlueCargo		Awaiting Pickup	Low	Box	2a84fe71-3b00-4
6	S1004	Richardview	Lake Micheleberg	Apparel	164.77	50.11	09-07-2025	11-07-2025	FastTrack	1567.37	Delayed	Medium	Container	3fecfa65b-0b22-4
7	S1005	South Zacharyland	Richardview	Books	3232.41	23.62	23-07-2025	31-07-2025	FastTrack	4062.11	Delayed	High	Pallet	f54a3672-abde-4
8	S1006	Robbinsville	Port Robin	Electronics	1266.9	97.51	04-08-2025	09-08-2025	EcoFreight	2372	Delivered	Low	Crate	956b68e9-a3ef-4
9	S1007	Richardview	New Yeseniaiville	Electronics	4295	9.91	17-08-2025	30-08-2025	BlueCargo	6674.6	In Transit	Medium	Crate	1ce9768a-bf39-4
10	S1008	Crystalshire	South Zacharyland	Food	3856.92		29-07-2025	10-08-2025	RapidTrans		Delayed	High	Container	9b691e60-e7ab-4
11	S1009	Jonathanstad	Port Robin	Apparel	1058.9	24.44	23-08-2025	03-09-2025	CargoIndia	1941.3	Awaiting Pickup	High	Pallet	4003dd0b-f7e3-4
12	S1010	Port Robin	West Jeffrey	Automotive	758.24	31.6	08-07-2025	13-07-2025	RapidTrans	1827.74	Awaiting Pickup	Low	Crate	3a47f952-0f05-40
13	S1011	Lake Micheleberg	Josephton	Food	1416.35	2.79	15-07-2025	29-07-2025	EcoFreight	2390.25	Awaiting Pickup	High	Container	8d1df874-08d0-4
14	S1012	Jonathanstad	Rogersville	Electronics	4688.43	57.24	20-07-2025	26-07-2025	RapidTrans	6539.83	Delayed	Medium	Crate	70ae9cd0-d850-4
15	S1013	South Zacharyland	Robbinsville	Chemicals	1749.63	16.52	25-08-2025	27-08-2025	FastTrack	3074.33	Delivered	Medium	Pallet	b07b21f7-0ff0-42
16	S1014	Richardborough	Port Jenniferhaven	Food	1149.3		13-07-2025	25-07-2025	RapidTrans		Delivered	Medium	Crate	bf1f480d-148b-40
17	S1015	Richardborough	Donnachester	Electronics	2183.63	78.35	15-08-2025	19-08-2025	CargoIndia	5126.63	Delivered	Low	Pallet	3881818d-1dba-4
18	S1016	Port Jenniferhaven	Richardborough	Furniture	4722.32	10.17	30-07-2025	03-08-2025	FastTrack	6748.52	In Transit	Medium	Pallet	30c33b8c-552d-4
19	S1017	Richardborough	Donnachester	Furniture	517.66	65.59	31-07-2025	05-08-2025	CargoIndia	1354.56	Delayed	Medium	Crate	c6cf0b5f-1e97-49
20	S1018	Crystalshire	West Marieland	Chemicals		86.66	04-07-2025	17-07-2025	EcoFreight		Awaiting Pickup	Medium	Pallet	30089030-0b51-4
21	S1019	Josephton	Brownshire	Machinery	1833.2	91.41	03-07-2025	12-07-2025	CargoIndia	2986.3	Delivered	High	Box	9bfa0da3-58bc-4
22	S1020	Port Robin	New Yeseniaiville	Food	1452.3	61.44	03-07-2025	11-07-2025	BlueCargo	4504.2	In Transit	Low	Pallet	eb6cb0d1-3c57-4
23	S1021	South Zacharyland	Bakermouth	Machinery	2372.45	49.56	03-08-2025	15-08-2025	FastTrack	5119.05	Awaiting Pickup	High	Pallet	f1e3cb9d-251a-4
24	S1022	Robbinsville	Stephensmouth	Apparel	2691.31	21.29	29-06-2025	13-07-2025	RapidTrans	3588.21	Delayed	High	Container	ac79826b-4764-4
25	S1023	Crystalshire	Jonathanstad	Chemicals	1665.81	64.74	07-07-2025	18-07-2025	BlueCargo	3154.21	In Transit	Medium	Crate	e8d00315-4627-4
26	S1024	Buchananchester	Port Jenniferhaven	Books	3806.01	58.32	05-08-2025	16-08-2025	RapidTrans	6302.71	Delivered	Low	Pallet	2d724270-abd7-4
27	S1025	Nancymouth	South Zacharyland	Electronics	1894.35	11.06	17-07-2025	24-07-2025	BlueCargo	4499.95	Delivered	Low	Box	7c1074b1-070a-4
28	S1026	Nancymouth	Buchananchester	Machinery	207.59	6.94	11-07-2025	21-07-2025	BlueCargo	1358.49	Awaiting Pickup	Low	Box	74544d48-a8db-4
29	S1027	Michaelshire	Michaelshire	Furniture	1120.84	45.3	22-08-2025	28-08-2025	BlueCargo	2166.34	Awaiting Pickup	Medium	Pallet	397ca51a-9547-4
30	S1028	Richardborough	Brownshire	Pharmaceuticals	2531.07	40.55	03-08-2025	17-08-2025	BlueCargo	3557.07	Delayed	Medium	Crate	95269229-8ec6-4
31	S1029	Robbinsville	Gomezton	Pharmaceuticals	1592.38	5.72	24-07-2025	28-07-2025	FastTrack	3228.08	Delivered	High	Box	5bc714e0-8bb6-4
32	S1030	Port Robin	South Zacharyland	Machinery	3285.64	92.31	18-07-2025	20-07-2025	EcoFreight	6649.24	Delivered	Medium	Crate	03f1e396-3055-4
33	S1031	Nancymouth	Buchananchester	Food	4852.89	80.18	17-08-2025	29-08-2025	RapidTrans	5727.69	In Transit	High	Pallet	a68d42b3-8f81-4
34	S1032	Port Jenniferhaven	Port Jenniferhaven	Machinery	4054.03	39.37	12-08-2025	15-08-2025	CargoIndia	6411.73	Delayed	High	Crate	b9a67d29-2be5-4
35	S1033	Jonathanstad	Crystalshire	Apparel	821.49	68.27	13-08-2025	21-08-2025	CargoIndia		Delivered	Medium	Pallet	5bcf3012-7241-4
36	S1034	West Jeffrey	Richardview	Electronics	600.71	26.47	22-08-2025	27-08-2025	EcoFreight	1015.41	In Transit	Low	Pallet	714b12-006c-4

What we do

- **Dataset Inspection**
 - Checked dataset shape (rows & columns)
 - Understood feature types (categorical, numerical, text)
- **Handling Missing Values**
 - Performed null value analysis
 - Imputed missing data using KNN Imputer for accurate filling
- **Data Scaling & Normalization**
 - Applied feature scaling to bring variables on the same scale
 - Ensured improved performance for ML algorithms
- **Final Preprocessed Dataset**
 - Cleaned, complete, and ready for model building
 - Balanced representation of features after transformation

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8379 entries, 0 to 8378
Data columns (total 31 columns):
```

#	Column	Non-Null Count	Dtype
0	Shipment_ID	8379	non-null
1	Origin	7728	non-null
2	Destination	7744	non-null
3	Cargo_Type	7754	non-null
4	Weight_kg	7539	non-null
5	Volume_m3	7631	non-null
6	Departure_Date	7733	non-null
7	Arrival_Date	7733	non-null
8	Carrier	7727	non-null
9	Cost_USD	6352	non-null
10	Status	7766	non-null
11	Priority	7763	non-null
12	Packaging_Type	7757	non-null
13	Customer_ID	7743	non-null
14	Tracking_Number	7755	non-null
15	Warehouse_Location	7745	non-null
16	Transit_Time_Days	7718	non-null
17	Distance_km	7718	non-null
18	Num_Checkpoints	7713	non-null
19	Temperature_C	7722	non-null
20	Humidity_Percent	7735	non-null
21	Delay_Hours	8379	non-null
22	Fuel_Consumed_L	7646	non-null
23	CO2_Emissions_kg	7646	non-null
24	Handling_Events	7725	non-null
25	Insurance_Value_USD	7758	non-null
26	Cargo_Value_USD	7693	non-null
27	Customs_Fees_USD	7730	non-null
28	No_of_Items	7704	non-null
29	Vibration_Events	7714	non-null
30	Transport_Method	8379	non-null

dtypes: float64(18), object(13)

memory usage: 2.0+ MB

Origin	651
Destination	635
Cargo_Type	625
Weight_kg	840
Volume_m3	748
Departure_Date	646
Arrival_Date	646
Carrier	652
Cost_USD	2027
Status	613
Priority	616
Packaging_Type	622
Customer_ID	636
Tracking_Number	624
Warehouse_Location	634
Transit_Time_Days	661
Distance_km	661
Num_Checkpoints	666
Temperature_C	657
Humidity_Percent	644
Delay_Hours	0
Fuel_Consumed_L	733
CO2_Emissions_kg	733
Handling_Events	654
Insurance_Value_USD	621
Cargo_Value_USD	686
Customs_Fees_USD	649
No_of_Items	675
Vibration_Events	665

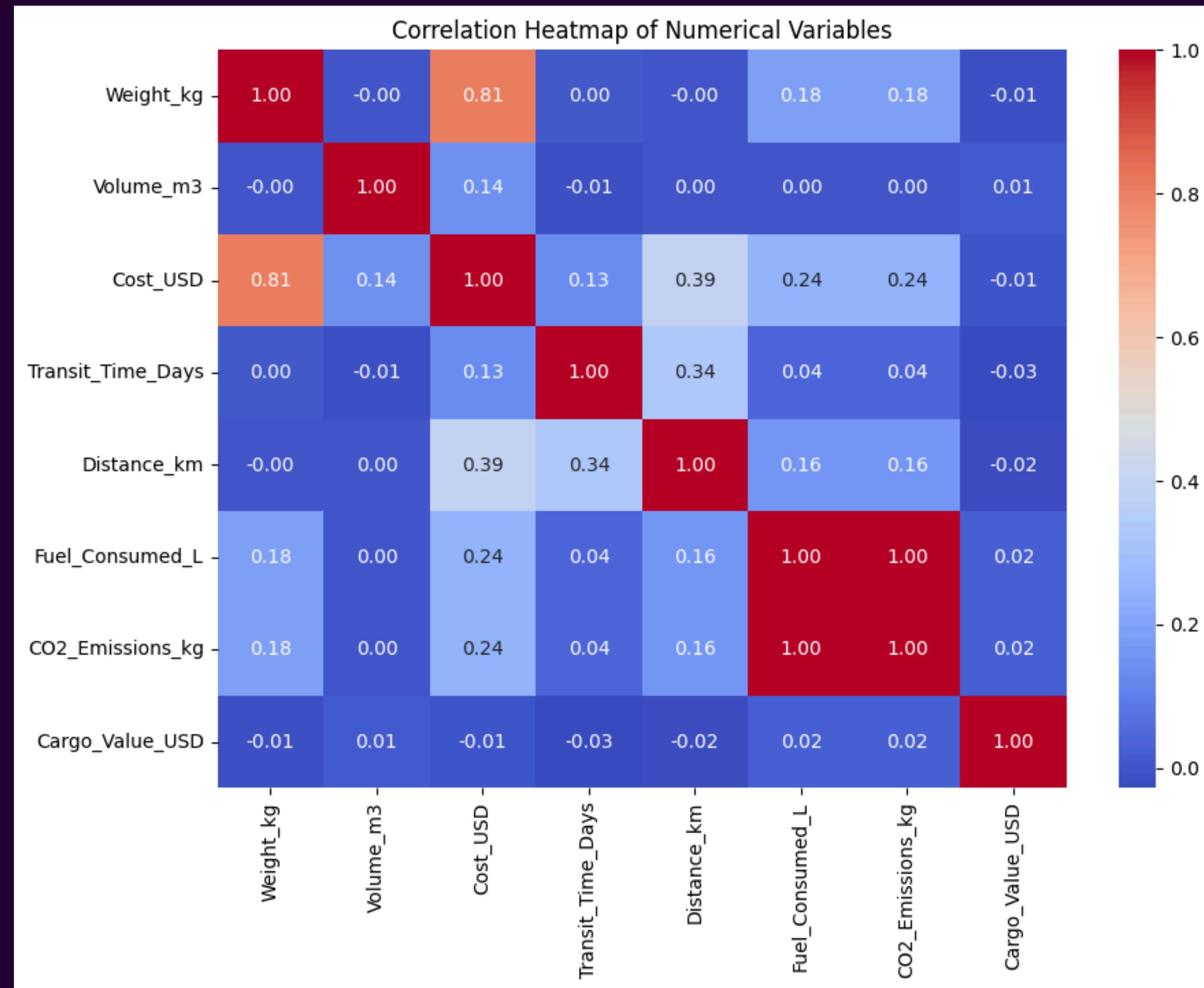
Before

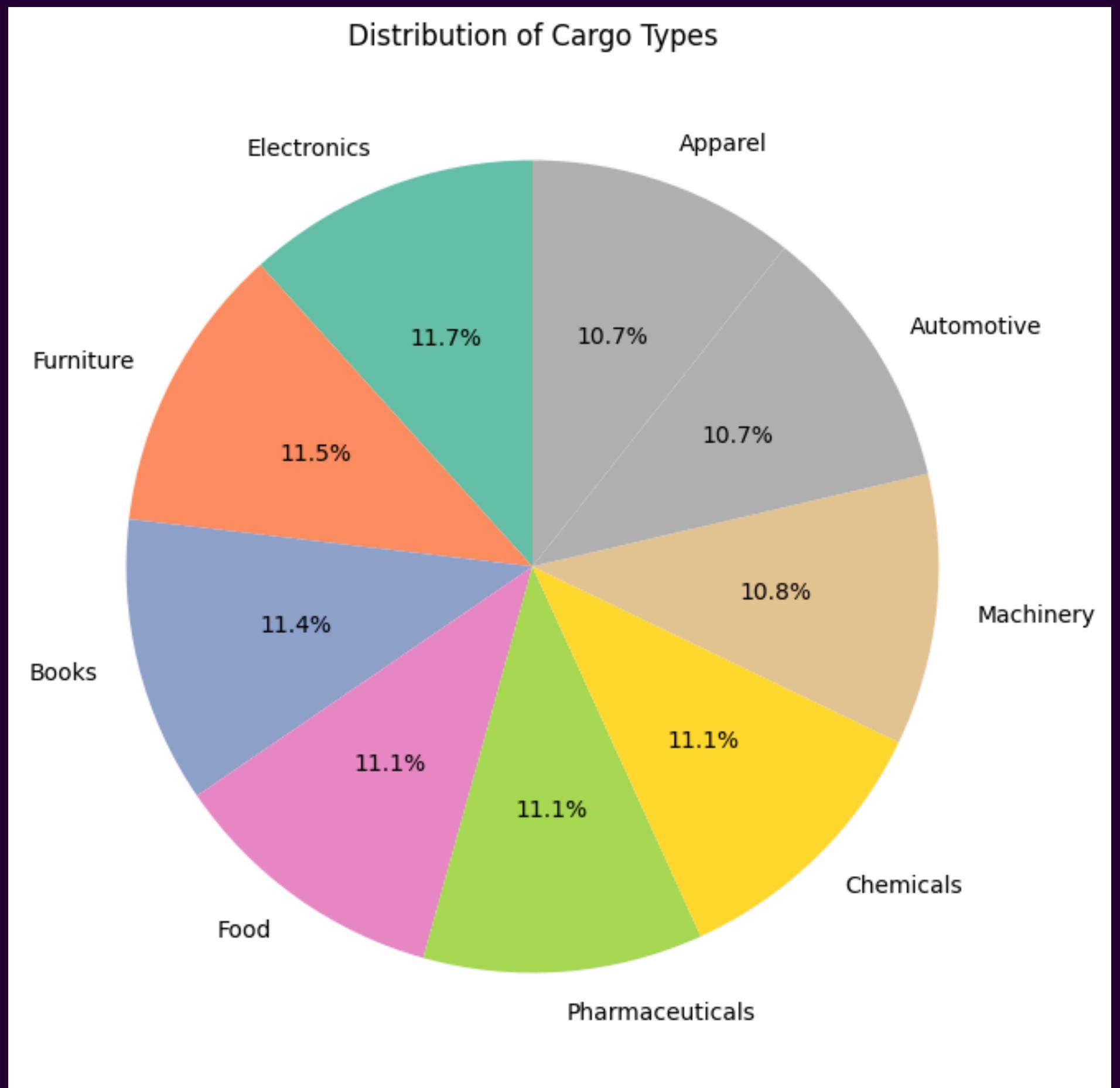
After

After Preprocessing

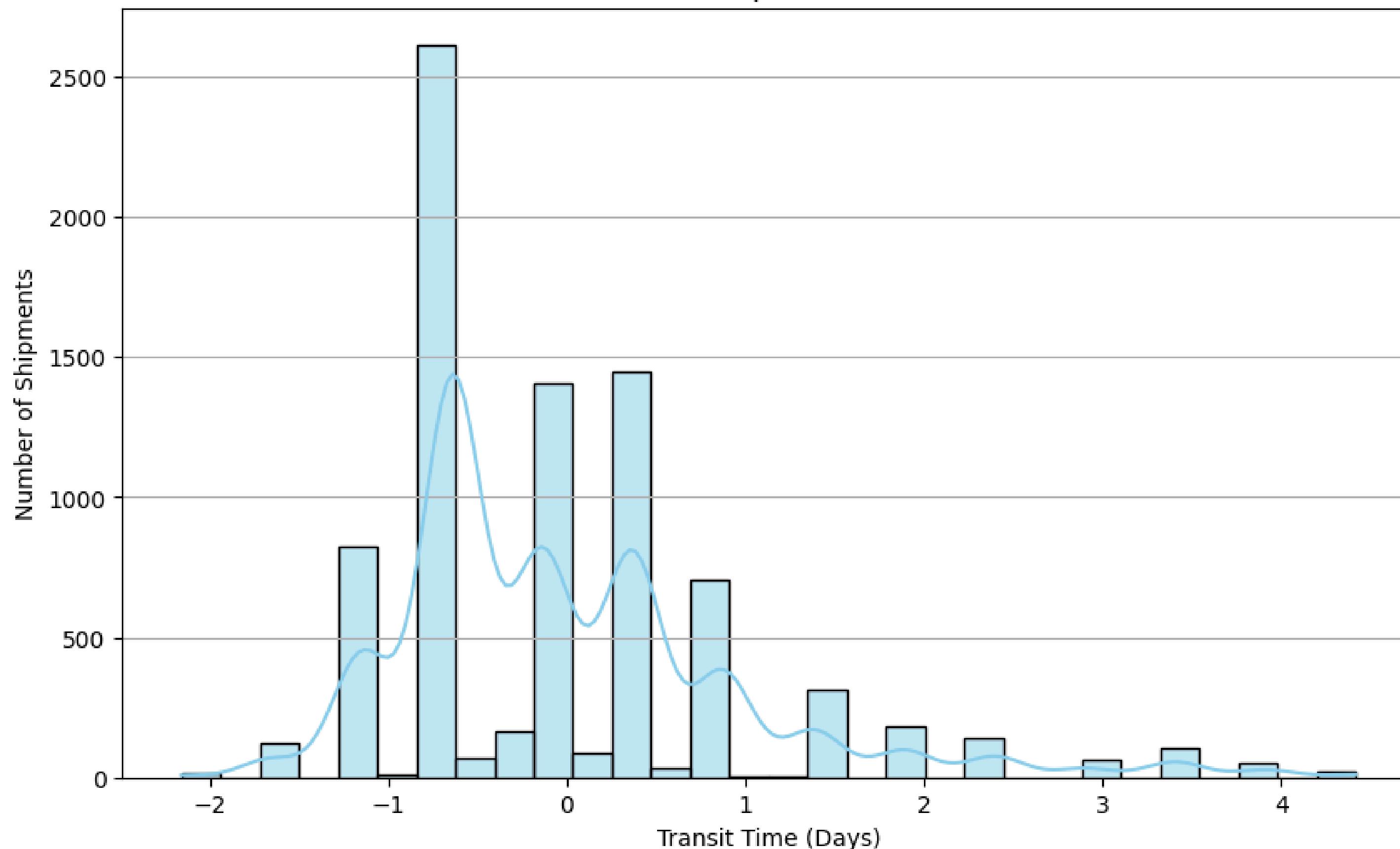
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Shipment_ID	Origin	Destination	Cargo_Type	Weight_kg	Volume_m3	Departure_Dat	Arrival_Date	Carrier	Cost_USD	Status	Priority	Packaging_Typ	Customer_ID	Tracking_Num	Warehouse_Lo	Transit_Time
2	S1000	Port Christopher	Stephensmout	Machinery	2611.87	42.86	30-06-2025	11-07-2025	RapidTrans	3723.47	In Transit	Low	Pallet	b67e9135-1bc	4.8537E+12	Sextonview	
3	S1001	Richardboroug	Gomezton	Furniture	2248.51	63.46	08-07-2025	16-07-2025	FastTrack	3383.11	Delayed	Medium	Container	b15ddc3b-4ec	9.57839E+12	Perkinschester	
4	S1002	West Jeffrey	Josephton	Pharmaceutical	1675.53	14.03	27-07-2025	30-07-2025	CargoIndia	3197.83	Delivered	Medium	Box	b8f15326-dbce	1.11507E+12	Martinside	
5	S1003	Port Jenniferha	Michaelshire	Machinery	2179.201667	85.98	17-08-2025	22-08-2025	BlueCargo	4461.516667	Awaiting Pickup	Low	Box	2a84fe71-3b00	3.87783E+12	Port Ronaldpor	
6	S1004	Richardview	Lake Michelebe	Apparel	164.77	50.11	09-07-2025	11-07-2025	FastTrack	1567.37	Delayed	Medium	Container	3fecab5b-0b22	5.54971E+12	New Nicoleview	
7	S1005	South Zacharyl	Richardview	Books	3232.41	23.62	23-07-2025	31-07-2025	FastTrack	4062.11	Delayed	High	Pallet	f54a3672-abde	6.23617E+11	Angelaport	
8	S1006	Robbinsville	Port Robin	Electronics	1266.9	97.51	04-08-2025	09-08-2025	EcoFreight	2372	Delivered	Low	Crate	956b68e9-a3e	8.15978E+12	West Christine	
9	S1007	Richardview	New Yeseniaiavl	Electronics	4295	9.91	17-08-2025	30-08-2025	BlueCargo	6674.6	In Transit	Medium	Crate	1ce9768a-bf39	4.07833E+12	West Tamara	1
10	S1008	Crystalshire	South Zacharyl	Food	3856.92	77.62	29-07-2025	10-08-2025	RapidTrans	4830.838333	Delayed	High	Container	9b691e60-e7a	2.7776E+12	West Mariah	
11	S1009	Jonathanstad	Port Robin	Apparel	1058.9	24.44	23-08-2025	03-09-2025	CargoIndia	1941.3	Awaiting Pickup	High	Pallet	4003dd0b-f7e3	8.7496E+12	New Dylan	
12	S1010	Port Robin	West Jeffrey	Automotive	758.24	31.6	08-07-2025	13-07-2025	RapidTrans	1827.74	Awaiting Pickup	Low	Crate	3a47f952-0f05	3.07576E+12	Lawrencestad	
13	S1011	Lake Michelebe	Josephton	Food	1416.35	2.79	15-07-2025	29-07-2025	EcoFreight	2390.25	Awaiting Pickup	High	Container	8d1df874-08d0	5.32748E+12	New Deborah	
14	S1012	Jonathanstad	Rogersville	Electronics	4688.43	57.24	20-07-2025	26-07-2025	RapidTrans	6539.83	Delayed	Medium	Crate	70ae9cd0-d85	9.80986E+11	Craigmouth	
15	S1013	South Zacharyl	Robbinsville	Chemicals	1749.63	16.52	25-08-2025	27-08-2025	FastTrack	3074.33	Delivered	Medium	Pallet	b07b21f7-0ff0-	9.90163E+12	East Jenniferbu	
16	S1014	Richardboroug	Port Jenniferha	Food	1149.3	45.20666667	13-07-2025	25-07-2025	RapidTrans	5362.708333	Delivered	Medium	Crate	bf1f480d-148b	4.89848E+11	Lisamouth	
17	S1015	Richardboroug	Donnachester	Electronics	2183.63	78.35	15-08-2025	19-08-2025	CargoIndia	5126.63	Delivered	Low	Pallet	3881818d-1db	1.44233E+12	Edgarview	
18	S1016	Port Jenniferha	Richardboroug	Furniture	4722.32	10.17	30-07-2025	03-08-2025	FastTrack	6748.52	In Transit	Medium	Pallet	30c33b8c-552	9.3421E+12	Port Dorothybu	
19	S1017	Richardboroug	Donnachester	Furniture	517.66	65.59	31-07-2025	05-08-2025	CargoIndia	1354.56	Delayed	Medium	Crate	c6cf0b5f-1e97	2.63565E+12	East Rickyton	
20	S1018	Crystalshire	West Marieland	Chemicals	2837.101667	86.66	04-07-2025	17-07-2025	EcoFreight	4243.081667	Awaiting Pickup	Medium	Pallet	30089030-0b5	7.84854E+12	Lake Lynnberg	
21	S1019	Josephton	Brownshire	Machinery	1833.2	91.41	03-07-2025	12-07-2025	CargoIndia	2986.3	Delivered	High	Box	9bfa0da3-58bc	8.87881E+12	North Marcus	
22	S1020	Port Robin	New Yeseniaiavl	Food	1452.3	61.44	03-07-2025	11-07-2025	BlueCargo	4504.2	In Transit	Low	Pallet	eb6cb0d1-3c5	2.84363E+11	Port Adam	
23	S1021	South Zacharyl	Bakermouth	Machinery	2372.45	49.56	03-08-2025	15-08-2025	FastTrack	5119.05	Awaiting Pickup	High	Pallet	f1e3cb9d-251a	7.42259E+12	Westburgh	
24	S1022	Robbinsville	Stephensmout	Apparel	2691.31	21.29	29-06-2025	13-07-2025	RapidTrans	3588.21	Delayed	High	Container	ac79826b-476	4.82987E+12	North Stephanii	
25	S1023	Crystalshire	Jonathanstad	Chemicals	1665.81	64.74	07-07-2025	18-07-2025	BlueCargo	3154.21	In Transit	Medium	Crate	e8d00315-462	2.70899E+12	East Miranda	
26	S1024	Buchananches	Port Jenniferha	Books	3806.01	58.32	05-08-2025	16-08-2025	RapidTrans	6302.71	Delivered	Low	Pallet	2d724270-abd	5.04316E+11	Sanchezberg	
27	S1025	Nancymouth	South Zacharyl	Electronics	1894.35	11.06	17-07-2025	24-07-2025	BlueCargo	4499.95	Delivered	Low	Box	7c1074b1-070	8.80882E+12	Aaronbury	1
28	S1026	Nancymouth	Buchananches	Machinery	207.59	6.94	11-07-2025	21-07-2025	BlueCargo	1358.49	Awaiting Pickup	Low	Box	74544d48-a8d	4.92118E+12	North Emilyview	
29	S1027	Michaelshire	Michaelshire	Furniture	1120.84	45.3	22-08-2025	28-08-2025	BlueCargo	2166.34	Awaiting Pickup	Medium	Pallet	397ca51a-954	3.40676E+12	Lake Whitneyc	
30	S1028	Richardboroug	Brownshire	Pharmaceutical	2531.07	40.55	03-08-2025	17-08-2025	BlueCargo	3557.07	Delayed	Medium	Crate	95269229-8ec	1.94824E+11	Zacharyburgh	
31	S1029	Robbinsville	Gomezton	Pharmaceutical	1592.38	5.72	24-07-2025	28-07-2025	FastTrack	3228.08	Delivered	High	Box	5bc714e0-8bb	5.42092E+12	Port Debra	
32	S1030	Port Robin	South Zacharyl	Machinery	3285.64	92.31	18-07-2025	20-07-2025	EcoFreight	6649.24	Delivered	Medium	Crate	03f1e396-3055	8.32454E+12	West Edwardto	
33	S1031	New - - - - -	Brownshires	Food	4852.89	80.18	17-08-2025	20-08-2025	RapidTrans	5727.09	In Transit	High	Pallet	-C8-1421-2-8681	1.00822E+12	South Androuck	

Exploratory Data Analysis

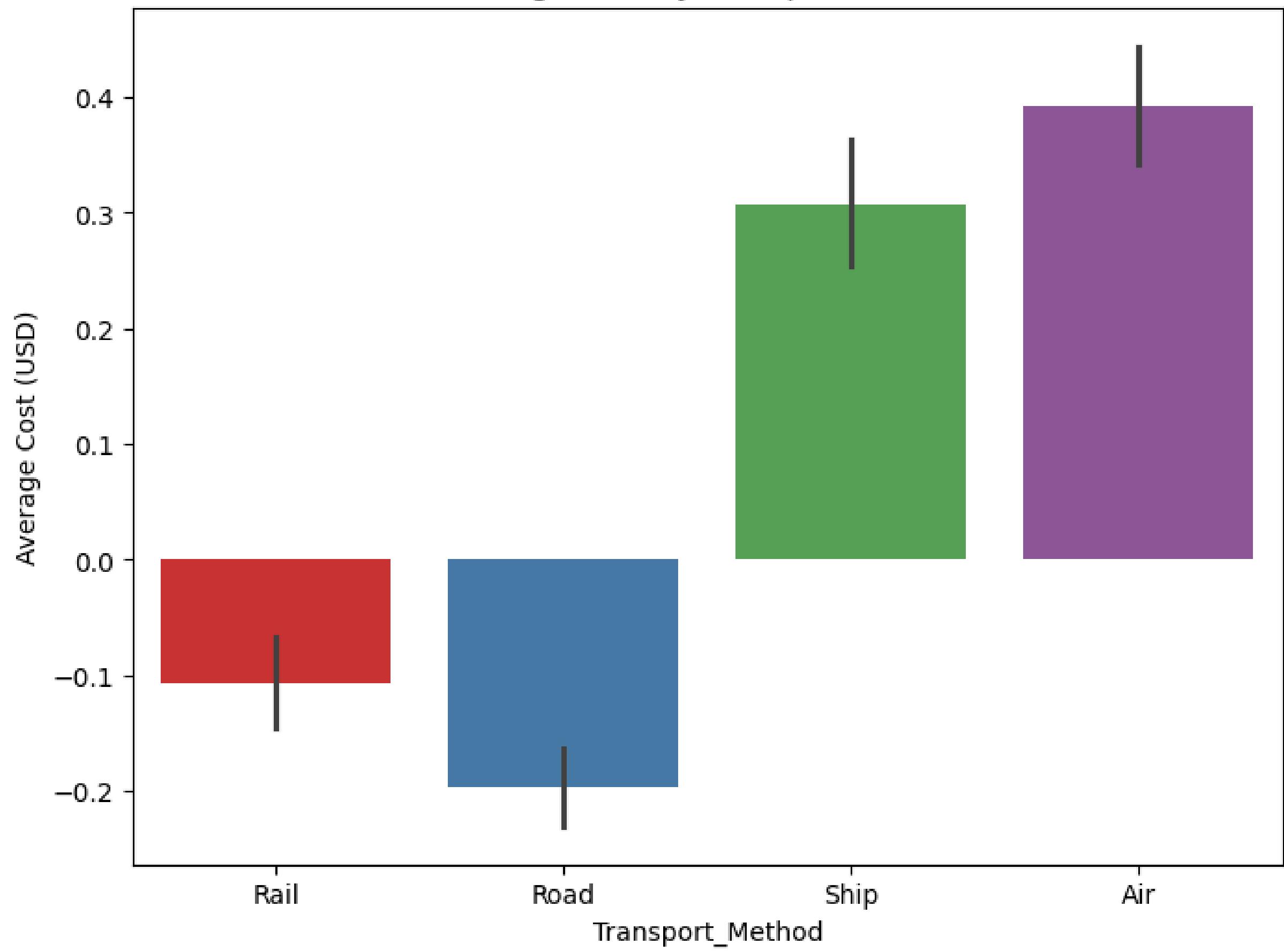




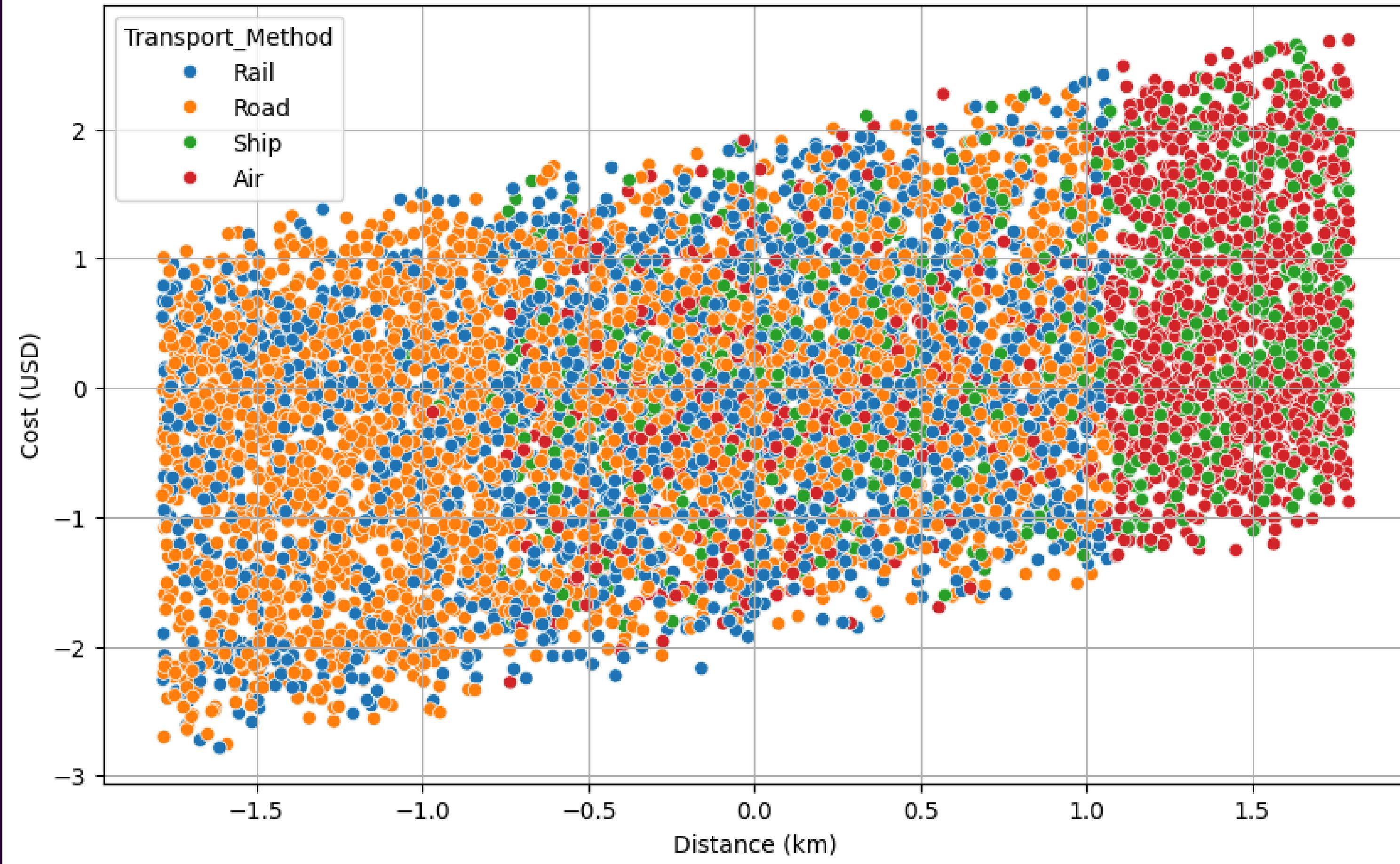
Distribution of Shipment Transit Times



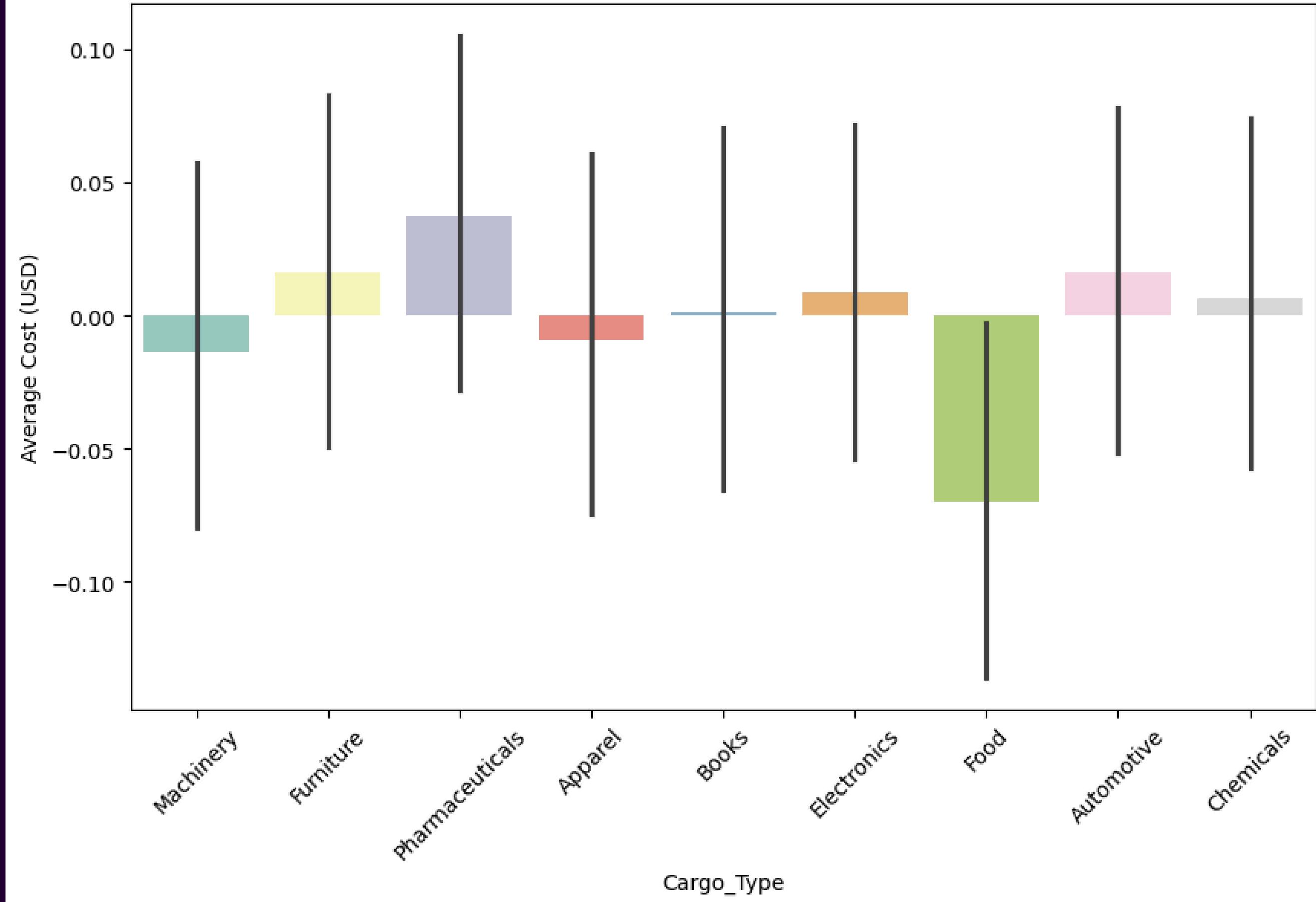
Average Cost by Transport Method



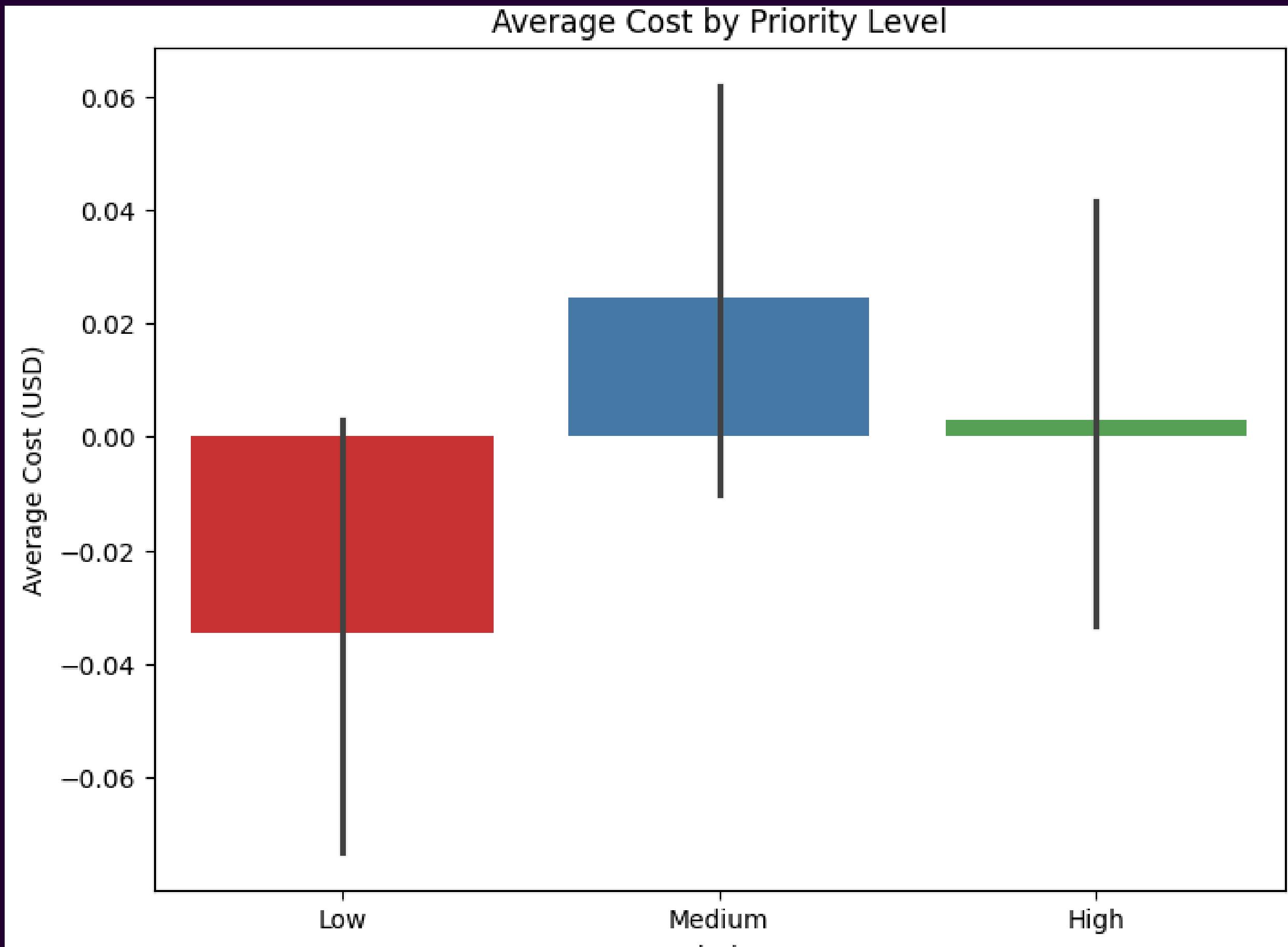
Shipment Cost vs. Distance by Transport Method



Average Cost by Cargo Type



Average Cost by Priority Level



CONCLUSION

- We finalized a clear problem statement addressing delays, costs, sustainability, and risks in logistics.
- Collected and explored a rich dataset covering trade, transport, and shipment characteristics.
- Understood the challenges in data quality such as missing values, outliers, and inconsistencies.
- Performed initial cleaning and preprocessing to prepare the dataset for analysis.
- Set the foundation for Phase 2, where we will dive into EDA, feature engineering, and actionable insights.

Thank you