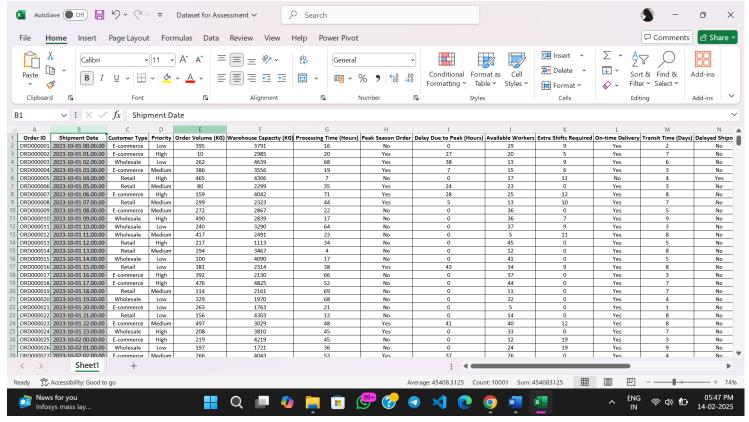# Section 1: Data Cleaning & Preparation (Excel & Google Sheets)

1. There are some inconsistencies in the dataset's Shipment Date format. Identify the inconsistencies, standardize the format, and show the before-and-after results.
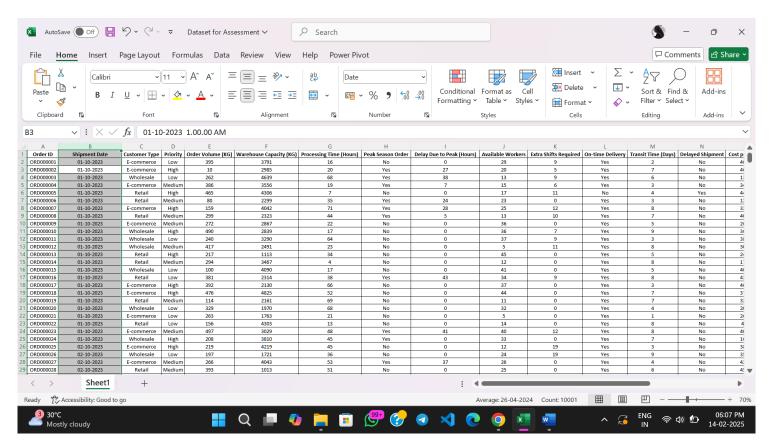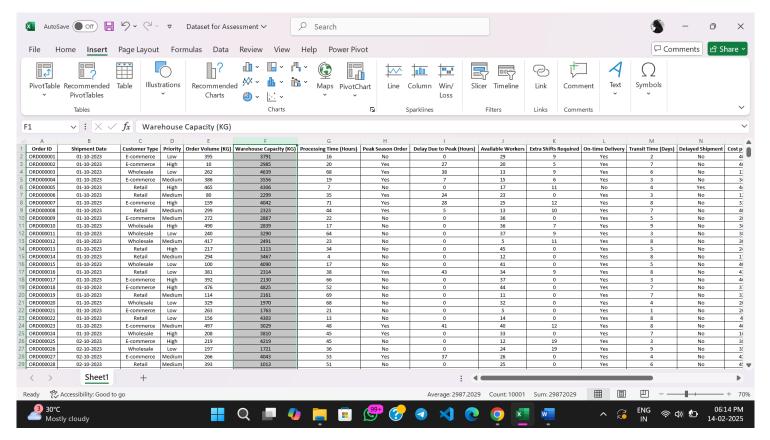
Before:

After:

Found inconsistencies in the **Shipment Date** format.
Standardized all dates to a common format (e.g., DD-MM-YYYY).

2.

The Warehouse Capacity (KG) column has some incorrect values due to manual entry errors. Write formulas in Excel or Google Sheets to detect and correct these anomalies based on Order Volume (KG).
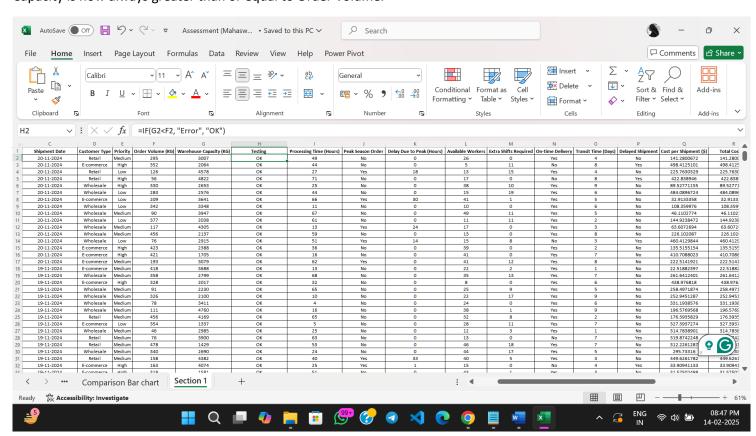
Before:

Order ID	Shipment Date	Customer Type	Priority	Order Volume (KG)	Warehouse Capacity (KG)	Processing Time (Hours)	Peak Season Order	Delay Due to Peak (Hours)	Available Workers	Extra Shifts Required	On-time Delivery	Transit Time (Days)	Delayed Shipment	Cost p
ORD000001	01-10-2023	E-commerce	Low	395	3791	16	No	0	29	9	Yes	2	No	4(
ORD000002	01-10-2023	E-commerce	High	10	2985	20	Yes	27	20	5	Yes	7	No	4(
ORD000003	01-10-2023	Wholesale	Low	262	4639	68	Yes	38	13	9	Yes	6	No	1:
ORD000004	01-10-2023	E-commerce	Medium	386	3556	19	Yes	7	15	6	Yes	3	No	3(
ORD000005	01-10-2023	Retail	High	465	4306	7	No	0	17	11	No	4	Yes	4(
ORD000006	01-10-2023	Retail	Medium	80	2299	35	Yes	24	23	0	Yes	3	No	1:
ORD000007	01-10-2023	E-commerce	High	159	4042	71	Yes	28	25	12	Yes	8	No	3:
ORD000008	01-10-2023	Retail	Medium	299	2323	44	Yes	5	13	10	Yes	7	No	4(
ORD000009	01-10-2023	E-commerce	Medium	272	2867	22	No	0	36	0	Yes	5	No	2(
ORD000010	01-10-2023	Wholesale	High	490	2839	17	No	0	36	7	Yes	9	No	3(
ORD000011	01-10-2023	Wholesale	Low	240	3290	64	No	0	37	9	Yes	3	No	3(
ORD000012	01-10-2023	Wholesale	Medium	417	2491	23	No	0	5	11	Yes	8	No	3(
ORD000013	01-10-2023	Retail	High	217	1113	34	No	0	45	0	Yes	5	No	2(
ORD000014	01-10-2023	Retail	Medium	294	3467	4	No	0	12	0	Yes	8	No	1:
ORD000015	01-10-2023	Wholesale	Low	100	4090	17	No	0	41	0	Yes	5	No	4(
ORD000016	01-10-2023	Retail	Low	381	2314	38	Yes	43	34	9	Yes	8	No	4:
ORD000017	01-10-2023	E-commerce	High	392	2130	66	No	0	37	0	Yes	3	No	4(
ORD000018	01-10-2023	E-commerce	High	476	4825	52	No	0	44	0	Yes	7	No	3:
ORD000019	01-10-2023	Retail	Medium	114	2161	69	No	0	11	0	Yes	7	No	3:
ORD000020	01-10-2023	Wholesale	Low	329	1970	68	No	0	32	0	Yes	4	No	2(
ORD000021	01-10-2023	E-commerce	Low	263	1763	21	No	0	5	0	Yes	1	No	2(
ORD000022	01-10-2023	Retail	Low	156	4303	13	No	0	14	0	Yes	8	No	4
ORD000023	01-10-2023	E-commerce	Medium	497	3029	48	Yes	41	40	12	Yes	8	No	4(
ORD000024	01-10-2023	Wholesale	High	208	3810	45	Yes	0	33	0	Yes	7	No	1(
ORD000025	02-10-2023	E-commerce	High	219	4219	45	No	0	12	19	Yes	3	No	3:
ORD000026	02-10-2023	Wholesale	Low	197	1721	36	No	0	24	19	Yes	9	No	3!
ORD000027	02-10-2023	E-commerce	Medium	266	4043	53	Yes	37	26	0	Yes	4	No	4:
ORD000028	02-10-2023	Retail	Medium	393	1013	51	No	0	25	0	Yes	6	No	4:

Average: 2987.2029 Count: 10001 Sum: 29872029
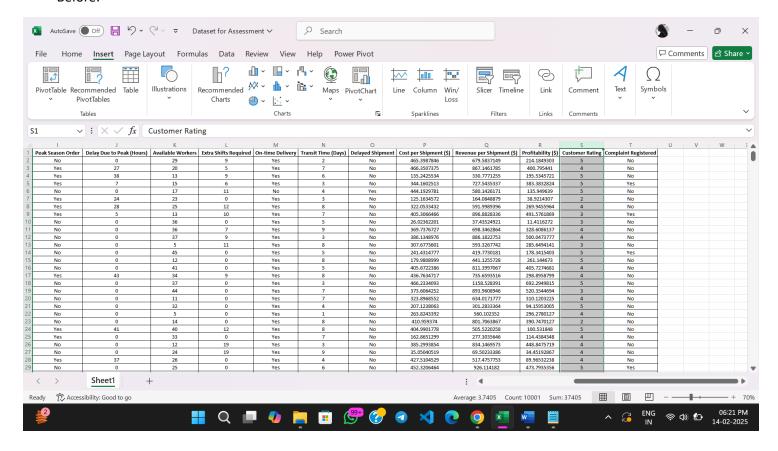
After:

Tested for incorrect values due to manual entry errors, created new column as "Testing" and noted as "OK"

Capacity is now always greater than or equal to Order Volume.

=IF(G2<F2, "Error", "OK")

Shipment Date	Customer Type	Priority	Order Volume (KG)	Warehouse Capacity (KG)	Testing	Processing Time (Hours)	Peak Season Order	Delay Due to Peak (Hours)	Available Workers	Extra Shifts Required	On-time Delivery	Transit Time (Days)	Delayed Shipment	Cost per Shipment ($)	Total Cos
20-11-2024	Retail	Medium	295	3007	OK	49	No	0	26	0	Yes	4	No	141.2800672	141.2800
20-11-2024	E-commerce	High	352	2064	OK	44	No	0	5	11	No	8	Yes	498.4125101	498.4125
20-11-2024	Retail	Low	126	4578	OK	27	Yes	18	13	15	Yes	4	No	225.7630329	225.7630
20-11-2024	Retail	High	56	4822	OK	71	No	0	17	0	No	8	Yes	422.838946	422.838
20-11-2024	Wholesale	High	330	2653	OK	25	No	0	38	10	Yes	9	No	89.52771155	89.52771
20-11-2024	Wholesale	Low	283	2576	OK	44	No	0	15	19	Yes	6	No	484.0896724	484.0896
20-11-2024	E-commerce	Low	209	3641	OK	66	Yes	30	41	1	Yes	5	No	32.9133358	32.9133
20-11-2024	Wholesale	Low	342	3348	OK	11	No	0	10	0	Yes	6	No	108.359976	108.359
20-11-2024	Wholesale	Medium	90	3947	OK	67	No	0	49	11	Yes	5	No	46.1102774	46.1102
20-11-2024	Wholesale	Low	377	2038	OK	61	No	0	11	11	Yes	2	No	144.9238472	144.9238
20-11-2024	Wholesale	Medium	117	4305	OK	13	Yes	24	17	0	Yes	3	No	63.6072694	63.6072
20-11-2024	Wholesale	Medium	456	2137	OK	59	No	0	13	0	Yes	8	No	226.102087	226.102
20-11-2024	Wholesale	Low	76	2915	OK	51	Yes	14	15	8	No	3	Yes	460.4129844	460.4129
20-11-2024	E-commerce	High	423	2388	OK	36	No	0	39	0	Yes	2	No	135.5155154	135.5155
20-11-2024	E-commerce	High	421	1705	OK	16	No	0	41	0	Yes	7	No	410.7088023	410.7088
20-11-2024	E-commerce	Medium	193	3079	OK	62	Yes	0	41	12	Yes	8	No	222.5141921	222.5141
19-11-2024	E-commerce	Medium	418	3688	OK	13	No	0	22	2	Yes	1	No	22.51882397	22.51882
19-11-2024	Wholesale	Medium	358	2799	OK	68	No	0	35	13	Yes	7	No	261.6412401	261.6412
19-11-2024	E-commerce	High	328	2017	OK	32	No	0	8	0	Yes	6	No	438.976818	438.976
19-11-2024	Wholesale	Medium	91	2230	OK	65	No	0	25	9	Yes	5	No	258.4971874	258.4971
19-11-2024	Wholesale	Medium	326	2100	OK	10	No	0	22	17	Yes	9	No	252.9451287	252.9451
19-11-2024	Wholesale	Medium	78	3411	OK	4	No	0	24	0	Yes	6	No	331.1938576	331.1938
19-11-2024	Wholesale	Medium	111	4760	OK	16	No	0	38	1	Yes	9	No	196.5769568	196.5769
19-11-2024	Retail	Medium	456	4169	OK	65	No	0	32	8	Yes	2	No	176.3935829	176.3935
19-11-2024	E-commerce	Low	354	1337	OK	5	No	0	28	11	Yes	7	No	327.3937274	327.3937
19-11-2024	Wholesale	Medium	46	2985	OK	25	No	0	12	3	Yes	1	No	314.7838901	314.7838
19-11-2024	Retail	Medium	76	3900	OK	63	No	0	13	0	No	7	Yes	319.8742248	42
19-11-2024	Retail	Medium	478	1429	OK	53	No	0	46	18	Yes	7	No	312.2261287	
19-11-2024	Wholesale	Medium	340	2690	OK	24	No	0	44	17	Yes	5	No	295.73316	63
19-11-2024	Retail	Medium	158	4382	OK	40	Yes	33	40	5	Yes	3	No	449.6261782	449.6261
19-11-2024	E-commerce	High	163	4074	OK	25	Yes	1	15	0	No	4	Yes	33.90941133	33.9094

3. Some Customer Rating values are missing. Use logical imputation techniques (e.g., median, mode, or weighted average based on Customer Type) to fill in the missing values and explain your approach.
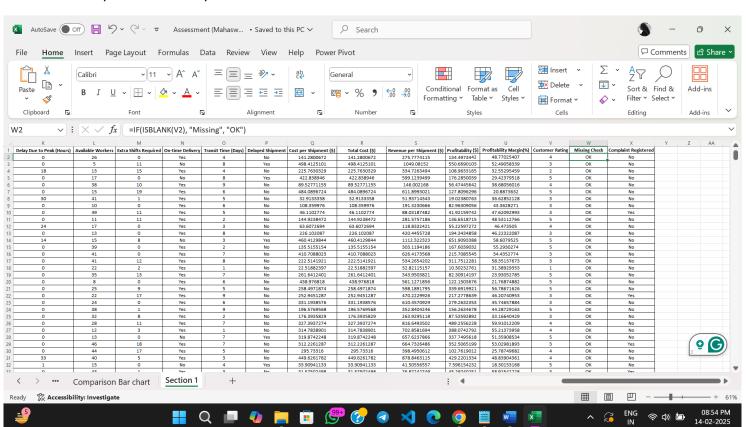
Before:

S1 — Customer Rating

Peak Season Order	Delay Due to Peak (Hours)	Available Workers	Extra Shifts Required	On-time Delivery	Transit Time (Days)	Delayed Shipment	Cost per Shipment ($)	Revenue per Shipment ($)	Profitability ($)	Customer Rating	Complaint Registered
No	0	29	9	Yes	2	No	465.3987846	679.5837149	214.1849303	5	No
Yes	27	20	5	Yes	7	No	466.3507375	867.1461785	400.795441	4	No
Yes	38	13	9	Yes	6	No	135.2425534	330.7771255	195.5345721	5	No
Yes	7	15	6	Yes	3	No	344.1602513	727.5435337	383.3832824	5	Yes
No	0	17	11	No	4	Yes	444.1929781	580.1426171	135.949639	5	No
Yes	24	23	0	Yes	3	No	125.1634572	164.0848879	38.9214307	2	No
Yes	28	25	12	Yes	8	No	322.0533432	591.9989396	269.9455964	4	No
Yes	5	13	10	Yes	7	No	405.3066466	896.8828336	491.5761869	3	Yes
No	0	36	0	Yes	5	No	26.02362201	37.43524921	11.4116272	3	No
No	0	36	7	Yes	9	No	369.7376727	698.3462864	328.6086137	4	No
No	0	37	9	Yes	3	No	386.1348976	886.1822753	500.0473777	4	No
No	0	5	11	Yes	8	No	307.6773601	593.3267742	285.6494141	3	No
No	0	45	0	Yes	5	No	241.4314777	419.7730181	178.3415403	5	Yes
No	0	12	0	Yes	8	No	179.9808999	441.1255728	261.144673	5	No
No	0	41	0	Yes	5	No	405.6722386	811.3997067	405.72774681	4	No
Yes	43	34	9	Yes	8	No	436.7634717	735.6593516	298.8958799	4	No
No	0	37	0	Yes	3	No	466.2334093	1158.528391	692.2949815	5	No
No	0	44	0	Yes	7	No	373.6064252	893.9608946	520.3544694	3	No
No	0	11	0	Yes	7	No	323.8968552	634.0171777	310.1203225	4	No
No	0	32	0	Yes	4	No	207.1238063	301.2833364	94.15953005	5	No
No	0	5	0	Yes	1	No	263.8243392	560.102352	296.2780127	4	No
No	0	14	0	Yes	8	No	410.959374	801.7063867	390.7470127	2	No
Yes	41	40	12	Yes	8	No	404.9901778	505.5220258	100.531848	5	No
Yes	0	33	0	Yes	7	No	162.8651299	277.3035646	114.4384348	4	No
No	0	12	19	Yes	3	No	385.2993854	834.1469573	448.8475719	4	No
No	0	24	19	Yes	9	No	35.05040519	69.50233386	34.45192867	4	No
Yes	37	26	0	Yes	4	No	427.5104529	517.4757753	89.96532238	4	No
No	0	25	0	Yes	6	No	452.3206464	926.114182	473.7935356	5	Yes

Sheet1

Average: 3.7405   Count: 10001   Sum: 37405

After:

Checked if any missing values in costumer rating or not.

no missing values are here, checked and created new column as "Missing Check" and noted as "OK"

Dataset is ready for further analysis

W2 — =IF(ISBLANK(V2), "Missing", "OK")

Delay Due to Peak (Hours)	Available Workers	Extra Shifts Required	On-time Delivery	Transit Time (Days)	Delayed Shipment	Cost per Shipment ($)	Total Cost ($)	Revenue per Shipment ($)	Profitability ($)	Profitability Margin(%)	Customer Rating	Missing Check	Complaint Registered
0	26	0	Yes	4	No	141.2800672	141.2800672	275.7774115	134.4973442	48.77025407	4	OK	No
0	5	11	No	8	Yes	498.4125101	498.4125101	1049.08152	550.6690103	52.49058339	3	OK	No
18	13	15	Yes	4	No	225.7630329	225.7630329	334.7263494	108.9633165	32.55295459	2	OK	No
0	17	0	No	8	Yes	422.838946	422.838946	599.1239499	176.2850039	29.42379518	5	OK	No
0	38	10	Yes	9	No	89.52771155	89.52771155	146.002168	56.47445642	38.68056016	4	OK	No
0	15	19	Yes	6	No	484.0896724	484.0896724	611.8993021	127.8096296	20.8873632	5	OK	No
30	41	1	Yes	5	No	32.9133358	32.9133358	51.93714343	19.02380763	36.62852128	3	OK	No
0	10	0	Yes	6	No	108.359976	108.359976	191.3230666	82.96309056	43.3628271	3	OK	No
0	49	11	Yes	5	No	46.1102774	46.1102774	88.03187482	41.92159742	47.62092993	3	OK	Yes
0	11	11	Yes	2	No	144.9238472	144.9238472	281.5757186	136.6518715	48.53112766	5	OK	No
24	17	0	Yes	3	No	63.6072694	63.6072694	118.8332421	55.22597272	46.473505	4	OK	No
0	13	0	Yes	8	No	226.102087	226.102087	420.4455728	194.3434858	46.22322087	3	OK	No
14	15	8	No	3	Yes	460.4129844	460.4129844	1112.322323	651.9093388	58.6079525	5	OK	No
0	39	0	Yes	2	No	135.5155154	135.5155154	303.1194186	167.6039032	55.2930274	5	OK	No
0	41	0	Yes	7	No	410.7088023	410.7088023	626.4173568	215.7085545	34.4352774	3	OK	No
0	41	12	Yes	8	No	222.5141921	222.5141921	534.2654202	311.7512281	58.35137673	3	OK	No
0	22	2	Yes	1	No	22.51882397	22.51882397	32.82115157	10.30232761	31.38929353	1	OK	No
0	35	13	Yes	7	No	261.6412401	261.6412401	343.9503821	82.30914197	23.93052785	5	OK	No
0	8	0	Yes	6	No	438.976818	438.976818	561.1271856	122.1503676	21.76874882	5	OK	No
0	25	9	Yes	5	No	258.4971874	258.4971874	598.1891795	339.6919921	56.78671626	5	OK	No
0	22	17	Yes	9	No	252.9451287	252.9451287	470.2229926	217.2778639	46.20740953	3	OK	Yes
0	24	0	Yes	6	No	331.1938576	331.1938576	610.4570929	279.2632353	45.74657884	3	OK	No
0	38	1	Yes	9	No	196.5769568	196.5769568	352.8404246	156.2634678	44.28729163	5	OK	No
0	32	8	Yes	2	No	176.3935829	176.3935829	263.9295118	87.53592892	33.16640429	3	OK	No
0	28	11	Yes	7	No	327.3937274	327.3937274	816.6493502	489.2556228	59.91012209	4	OK	No
0	12	3	Yes	1	No	314.7838901	314.7838901	702.8581694	388.0742792	55.21373958	4	OK	No
0	13	0	No	7	Yes	319.8742248	319.8742248	657.6237866	337.7495618	51.35908534	5	OK	No
0	46	18	Yes	7	No	312.2261287	312.2261287	664.7326486	352.5065199	53.02981893	3	OK	Yes
0	44	17	Yes	5	No	295.73316	295.73316	398.4950612	102.7619012	25.78749682	4	OK	No
33	40	5	Yes	3	No	449.6261782	449.6261782	878.8463115	429.2201334	48.83904361	4	OK	No
1	15	0	No	4	Yes	33.90941133	33.90941133	41.50556557	7.596154232	18.30153168	5	OK	No

Comparison Bar chart   Section 1

4. Identify **duplicate records** based on Order ID and remove them while preserving the latest entry.
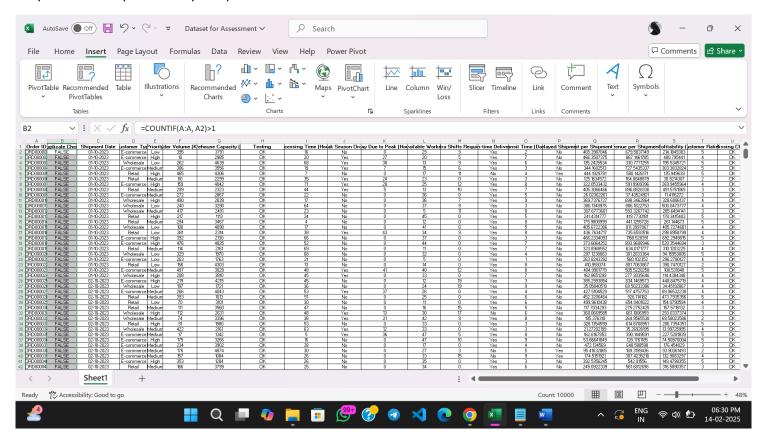
Before:

After:

Found 3 duplicate Order IDs. it was removed by using "countif" formula

All duplicate Order IDs are removed.

Only the latest shipment entry is kept.

Section 4: Problem-Solving & Decision Making:

13. Based on your analysis, suggest three strategic actions to reduce peak season delays.

    o **Increase Workforce & Shifts** – Some warehouses struggle more than others during peak times. Adding
    extra shifts or temporary staff in high-demand locations can help speed up processing.

    o **Prioritize Urgent Orders** – Not all orders need the same urgency. Orders from high-priority customers should be routed through faster processing or dedicated lanes.

    o **Improve Forecasting & Planning** – Using past data, the company can predict high-demand periods and prepare resources in advance, like stocking up inventory closer to customers.

14. If the company wants to increase profitability by 10%, recommend changes in cost, revenue, or processing strategies using data insights

    o **Cut Down on Expensive Shipments** – Some warehouses have higher costs per shipment. Optimizing which orders go through which warehouses can reduce unnecessary expenses.

    o **Introduce Surge Pricing** – During peak times, a small increase in shipping fees for urgent deliveries can help cover extra costs.

    o **Automate Repetitive Tasks** – Manual processes lead to delays and errors. Using automated sorting, tracking, and dispatching can save time and labor costs.

15. Identify a correlation between Transit Time (Days) and Customer Rating using a scatter plot in Tableau/Power BI. Interpret whether longer transit times impact customer ratings

    we use a scatter plot with:
    o **Transit Time (Days) on the X-axis**

    o **Customer Rating on the Y-axis**

    If the trend shows that longer delivery times lead to lower ratings, it's a sign that delays are frustrating customers. If the impact is small, other factors (like customer service or product quality) might be more important.