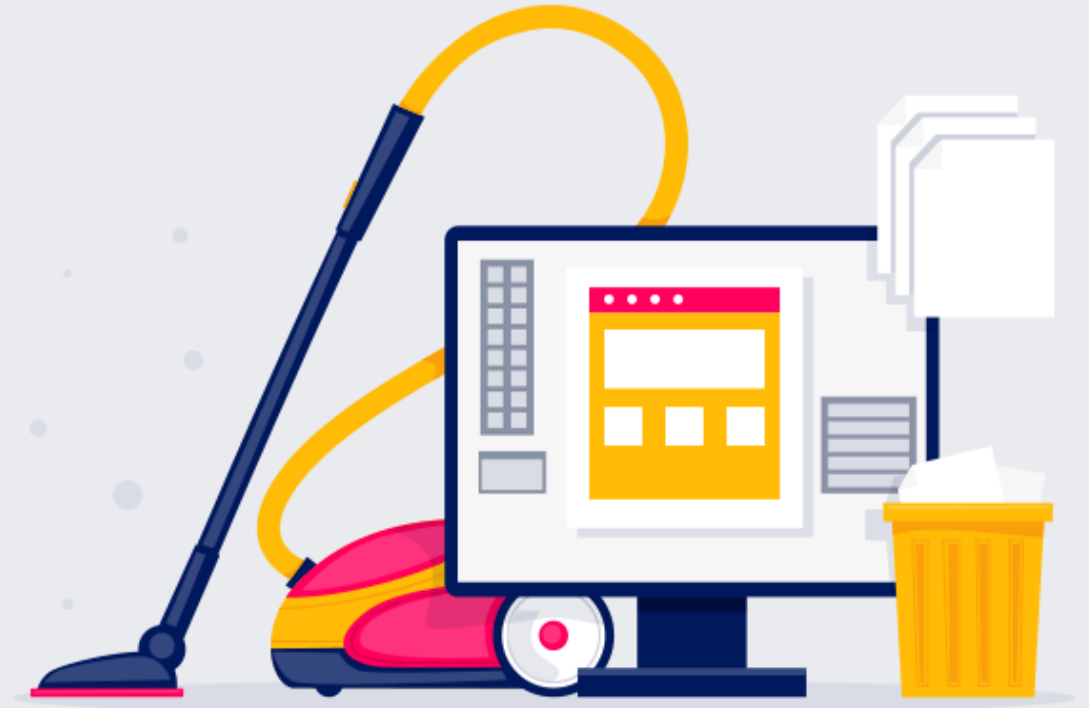




A step by step approach to structuring badly structured data in **Power Query**



Hi Olanrewaju,

You are welcome to the team, now is the time to put your excel skill to work.

One of our Clients just sent us sample invoices data for Analysis.
The client has requested that we put the data into proper shape and possibly assess the quality of the data.

Your task is to **Automate** the **Data Cleaning** and have this data in the proper rows and columns.

There will be a meeting with this client on Monday 19th December, 2022. Please kindly ensure that the cleaned data is available by then.

Attached to this email is the dataset:

<https://foresightbi.com.ng/wp-content/uploads/2020/05/1.-Badly-Structured-Sales-Data-1.xlsx>

Kind regards,
Daniel Ayangbile
Senior Analyst

Analyst: Olanrewaju Olatunji



#Let me walk you through how I cleaned the Dataset

From this

ABC 123 Column1	ABC 123 Column2	ABC 123 Column3	ABC 123 Column4	ABC 123 Column5	ABC 123 Column6	ABC 123 Column7	ABC 123 Column8
Segment>>	Consumer	null	null	null	Consumer Total	Corporate	
Ship Mode>>	First Class	Same Day	Second Class	Standard Class	null	First Class	Se
Order ID	null	null	null	null	null	null	
CA-2011-100293	null	null	null	null	null	null	
CA-2011-100706	null	null	129.44	null	129.44	null	
CA-2011-100895	null	null	null	605.47	605.47	null	
CA-2011-100916	null	null	null	null	null	null	
CA-2011-101266	null	null	13.36	null	13.36	null	
CA-2011-101560	null	null	null	null	null	null	
CA-2011-101770	null	null	null	null	null	null	
CA-2011-102274	null	null	null	null	null	null	
CA-2011-102673	null	null	null	null	null	null	
CA-2011-102988	null	null	null	null	null	null	
CA-2011-103317	null	null	null	null	null	242.546	
CA-2011-103366	149.95	null	null	null	149.95	null	
CA-2011-103807	null	null	null	null	null	null	
CA-2011-103989	null	null	null	null	null	590.762	
CA-2011-104283	null	null	null	616.14	616.14	null	
CA-2011-106054	null	null	null	null	null	12.78	
CA-2011-106810	null	null	null	null	null	null	
CA-2011-107573	null	null	null	23.472	23.472	null	

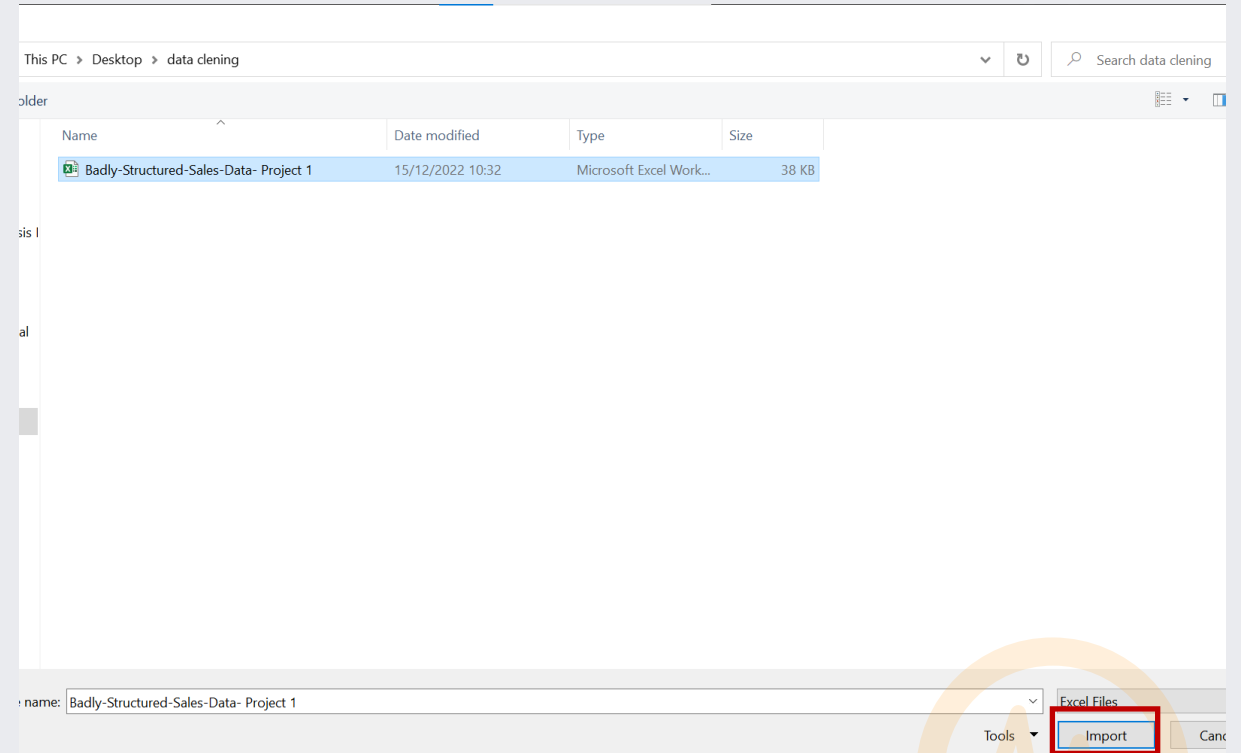
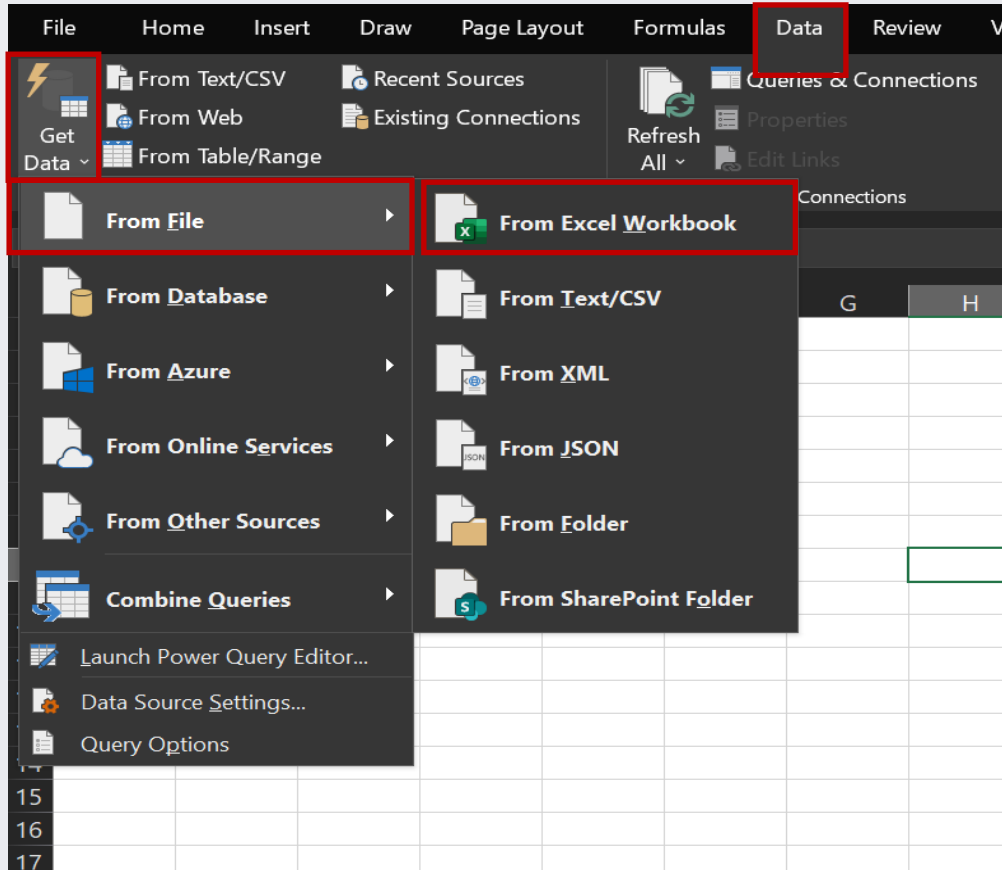


To this

ABC Order_Id	ABC Shipment_Mode	ABC Segment	\$ Sales Amount
CA-2011-103366	First Class	Consumer	149.95
CA-2011-109043	First Class	Consumer	243.60
CA-2011-113166	First Class	Consumer	9.57
CA-2011-124023	First Class	Consumer	8.96
CA-2011-130155	First Class	Consumer	34.20
CA-2011-136861	First Class	Consumer	31.98
CA-2011-153927	First Class	Consumer	286.65
CA-2011-157784	First Class	Consumer	514.03
CA-2011-160094	First Class	Consumer	1,000.95
CA-2011-164749	First Class	Consumer	9.91
CA-2011-166730	First Class	Consumer	39.13
CA-2012-102722	First Class	Consumer	106.50
CA-2012-102778	First Class	Consumer	18.18
CA-2012-117828	First Class	Consumer	194.32
CA-2012-130218	First Class	Consumer	59.48
CA-2012-132318	First Class	Consumer	182.91
CA-2012-137974	First Class	Consumer	2,298.90
CA-2012-138625	First Class	Consumer	197.72
CA-2012-141327	First Class	Consumer	440.14
CA-2012-149300	First Class	Consumer	32.99
CA-2012-150560	First Class	Consumer	196.62
CA-2012-165414	First Class	Consumer	47.98



Open up Excel workbook then navigate through;
Data → Get Data → From File → Excel Workbook to Import Data





Click on the **Dataset** and then **Transform Data**

Navigator

☐ Select multiple items

Display Options ▾

Badly-Structured-Sales-Data- Project 1.xlsx [1]

Dirty 1

Dirty 1

Segment>>	Consumer	Column3	Column4	Column5
Ship Mode>>	First Class	Same Day	Second Class	Standard Class
Order ID	null	null	null	
CA-2011-100293	null	null	null	
CA-2011-100706	null	null	129.44	
CA-2011-100895	null	null	null	60
CA-2011-100916	null	null	null	
CA-2011-101266	null	null	13.36	
CA-2011-101560	null	null	null	
CA-2011-101770	null	null	null	
CA-2011-102274	null	null	null	
CA-2011-102673	null	null	null	
CA-2011-102988	null	null	null	
CA-2011-103317	null	null	null	
CA-2011-103366	149.95	null	null	
CA-2011-103807	null	null	null	
CA-2011-103989	null	null	null	
CA-2011-104283	null	null	null	61
CA-2011-106054	null	null	null	
CA-2011-106810	null	null	null	
CA-2011-107573	null	null	null	23
CA-2011-107811	null	null	null	
CA-2011-108707	null	null	null	

Load ▾ **Transform Data** Cancel





Yah! Here we have the **Dirty Data**

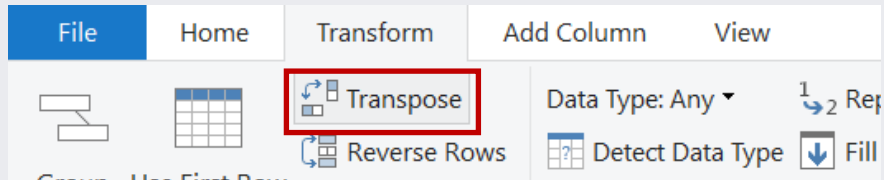
= Source{[Item="Dirty 1",Kind="Sheet"]}[Data]																						
ABC 123	Column1	ABC 123	Column2	ABC 123	Column3	ABC 123	Column4	ABC 123	Column5	ABC 123	Column6	ABC 123	Column7	ABC 123	Column8							
1	Segment>>	Consumer		null		null		null		Consumer Total		Corporate										
2	Ship Mode>>	First Class	Same Day		Second Class		Standard Class			null		First Class		Sam								
3	Order ID		null		null		null		null		null		null									
4	CA-2011-100293		null		null		null		null		null		null									
5	CA-2011-100706		null		null		129.44		null		129.44		null									
6	CA-2011-100895		null		null		null		605.47		605.47		null									
7	CA-2011-100916		null		null		null		null		null		null									
8	CA-2011-101266		null		null		13.36		null		13.36		null									
9	CA-2011-101560		null		null		null		null		null		null									
10	CA-2011-101770		null		null		null		null		null		null									
11	CA-2011-102274		null		null		null		null		null		null									
12	CA-2011-102673		null		null		null		null		null		null									
13	CA-2011-102988		null		null		null		null		null		null									
14	CA-2011-103317		null		null		null		null		null		242.546									
15	CA-2011-103366		149.95		null		null		null		149.95		null									
16	CA-2011-103807		null		null		null		null		null		null									
17	CA-2011-103989		null		null		null		null		null		590.762									
18	CA-2011-104283		null		null		null		616.14		616.14		null									
19	CA-2011-106054		null		null		null		null		null		12.78									
20	CA-2011-106810		null		null		null		null		null		null									
21	CA-2011-107573		null		null		null		23.472		23.472		null									
22																						





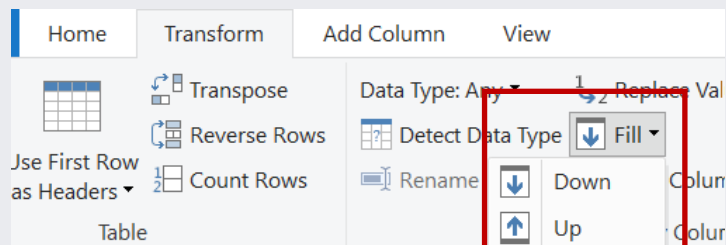
4

Select the first **Column** then **Transpose**



ABC 123 Column1	ABC 123 Column2	ABC 123 Column3	ABC 123 Column4	ABC 123 Column5	ABC 123 Column6	ABC 123 Column7	ABC 123 Column8
Segment>>	Ship Mode>>	Order ID	CA-2011-100293	CA-2011-100706	CA-2011-100895	CA-2011-100916	CA-2011-100916
Consumer	First Class	null	null	null	null	null	null
null	Same Day	null	null	null	null	null	null
null	Second Class	null	null	129.44	null	null	null
null	Standard Class	null	null	null	605.47	null	null
Consumer Total	null	null	null	129.44	605.47	null	null
Corporate	First Class	null	null	null	null	null	null
null	Same Day	null	null	null	null	null	null
null	Second Class	null	null	null	null	null	null
null	Standard Class	null	null	null	null	788.86	788.86
Corporate Total	null	null	null	null	null	788.86	788.86
Home Office	First Class	null	null	null	null	null	null
null	Same Day	null	null	null	null	null	null
null	Second Class	null	null	null	null	null	null
null	Standard Class	null	91.056	null	null	null	null
Home Office Total	null	null	91.056	null	null	null	null

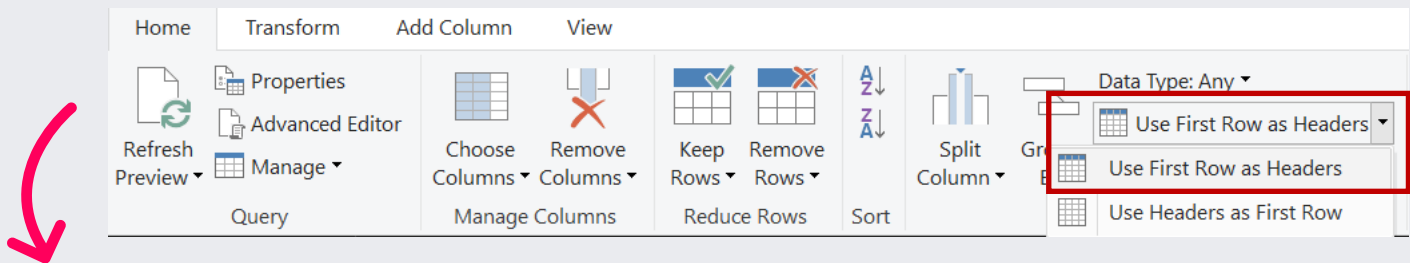
5 Select the first two Columns then Fill Down to replace null values



ABC 123 Column1	ABC 123 Column2	ABC 123 Column3	ABC 123 Column4	ABC 123 Column5	ABC 123 Column6	ABC 123 Column7	ABC 123 Column8
Segment>>	Ship Mode>>	Order ID	CA-2011-100293	CA-2011-100706	CA-2011-100895	CA-2011-100916	CA-2011-100917
Consumer	First Class	null	null	null	null	null	null
Consumer	Same Day	null	null	null	null	null	null
Consumer	Second Class	null	null	129.44	null	null	null
Consumer	Standard Class	null	null	null	605.47	null	null
Consumer Total	Standard Class	null	null	129.44	605.47	null	null
Corporate	First Class	null	null	null	null	null	null
Corporate	Same Day	null	null	null	null	null	null
Corporate	Second Class	null	null	null	null	null	null
Corporate	Standard Class	null	null	null	null	788.86	null
Corporate Total	Standard Class	null	null	null	null	788.86	null
Home Office	First Class	null	null	null	null	null	null
Home Office	Same Day	null	null	null	null	null	null
Home Office	Second Class	null	null	null	null	null	null
Home Office	Standard Class	null	91.056	null	null	null	null
Home Office Total	Standard Class	null	91.056	null	null	null	null




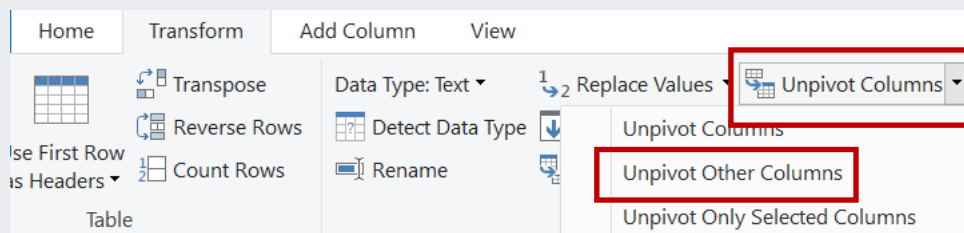
6 Use first Row as Header



ABC 123 Segment>>	ABC 123 Ship Mode>>	ABC 123 Order ID	ABC 123 CA-2011-100293	ABC 123 CA-2011-100706	ABC 123 CA-2011-100895
Consumer	First Class	null	null	null	null
Consumer	Same Day	null	null	null	null
Consumer	Second Class	null	null	129.44	null
Consumer	Standard Class	null	null	null	605.47
Consumer Total	Standard Class	null	null	129.44	605.47
Corporate	First Class	null	null	null	null
Corporate	Same Day	null	null	null	null
Corporate	Second Class	null	null	null	null
Corporate	Standard Class	null	null	null	null
Corporate Total	Standard Class	null	null	null	null
Home Office	First Class	null	null	null	null
Home Office	Same Day	null	null	null	null
Home Office	Second Class	null	null	null	null
Home Office	Standard Class	null	91.056	null	null
Home Office Total	Standard Class	null	91.056	null	null



7 Select the first two Columns then Unpivot other column



ABC 123	Segment>>	ABC 123	Ship Mode>>	ABC C	Attribute	ABC 123	Value
	Consumer		First Class		CA-2011-103366		149.95
	Consumer		First Class		CA-2011-109043		243.6
	Consumer		First Class		CA-2011-113166		9.568
	Consumer		First Class		CA-2011-124023		8.96
	Consumer		First Class		CA-2011-130155		34.2
	Consumer		First Class		CA-2011-136861		31.984
	Consumer		First Class		CA-2011-153927		286.65
	Consumer		First Class		CA-2011-157784		514.03
	Consumer		First Class		CA-2011-160094		1000.95
	Consumer		First Class		CA-2011-164749		9.912
	Consumer		First Class		CA-2011-166730		39.128
	Consumer		First Class		CA-2012-102722		106.5
	Consumer		First Class		CA-2012-102778		18.176
	Consumer		First Class		CA-2012-117828		194.32
	Consumer		First Class		CA-2012-130218		59.48
	Consumer		First Class		CA-2012-132318		182.91
	Consumer		First Class		CA-2012-137974		2298.9
	Consumer		First Class		CA-2012-138625		197.72
	Consumer		First Class		CA-2012-141327		440.144
	Consumer		First Class		CA-2012-149300		32.985
	Consumer		First Class		CA-2012-150560		196.62
	Consumer		First Class		CA-2012-165414		47.976



Hold down ctrl + select all the Columns,
right click then **Remove Duplicates**

ABC Merged.1	ABC Merged.2	ABC Attribute	ABC 123 Value
First Class	Consumer	CA-2011-103366	
First Class	Consumer	CA-2011-109043	
First Class	Consumer	CA-2011-113166	
First Class	Consumer	CA-2011-124023	
First Class	Consumer	CA-2011-130155	
First Class	Consumer	CA-2011-136861	
First Class	Consumer	CA-2011-153927	
First Class	Consumer	CA-2011-157784	
First Class	Consumer	CA-2011-160094	
First Class	Consumer	CA-2011-164749	
First Class	Consumer	CA-2011-166730	
First Class	Consumer	CA-2012-102722	
First Class	Consumer	CA-2012-102778	
First Class	Consumer	CA-2012-117828	
First Class	Consumer	CA-2012-130218	59.48
First Class	Consumer	CA-2012-132318	182.91
First Class	Consumer	CA-2012-137974	2298.9
First Class	Consumer	CA-2012-138625	197.72
First Class	Consumer	CA-2012-141327	440.144
First Class	Consumer	CA-2012-149300	32.985
First Class	Consumer	CA-2012-150560	196.62
First Class	Consumer	CA-2012-165414	47.976

Copy

Remove Columns

Remove Other Columns

Add Column From Examples...

Remove Duplicates

Remove Errors

Replace Values...

Fill

Change Type

Merge Columns

Group By...

Unpivot Columns

Unpivot Only Selected Columns

Move





9 Rename, Reorder, and Change each Column Type

AB_C Order_Id ▾	AB_C Shipment_Mode ▾	AB_C Segment ▾	\$ Sales Amount ▾
CA-2011-103366	First Class	Consumer	149.95
CA-2011-109043	First Class	Consumer	243.60
CA-2011-113166	First Class	Consumer	9.57
CA-2011-124023	First Class	Consumer	8.96
CA-2011-130155	First Class	Consumer	34.20
CA-2011-136861	First Class	Consumer	31.98
CA-2011-153927	First Class	Consumer	286.65
CA-2011-157784	First Class	Consumer	514.03
CA-2011-160094	First Class	Consumer	1,000.95
CA-2011-164749	First Class	Consumer	9.91
CA-2011-166730	First Class	Consumer	39.13
CA-2012-102722	First Class	Consumer	106.50
CA-2012-102778	First Class	Consumer	18.18
CA-2012-117828	First Class	Consumer	194.32
CA-2012-130218	First Class	Consumer	59.48
CA-2012-132318	First Class	Consumer	182.91
CA-2012-137974	First Class	Consumer	2,298.90
CA-2012-138625	First Class	Consumer	197.72
CA-2012-141327	First Class	Consumer	440.14
CA-2012-149300	First Class	Consumer	32.99
CA-2012-150560	First Class	Consumer	196.62
CA-2012-165414	First Class	Consumer	47.98





10 Clean Data!!!

Order_Id	Shipment_Mode	Segment	Sales Amount
CA-2011-103366	First Class	Consumer	149.95
CA-2011-109043	First Class	Consumer	243.6
CA-2011-113166	First Class	Consumer	9.568
CA-2011-124023	First Class	Consumer	8.96
CA-2011-130155	First Class	Consumer	34.2
CA-2011-136861	First Class	Consumer	31.984
CA-2011-153927	First Class	Consumer	286.65
CA-2011-157784	First Class	Consumer	514.03
CA-2011-160094	First Class	Consumer	1000.95
CA-2011-164749	First Class	Consumer	9.912
CA-2011-166730	First Class	Consumer	39.128
CA-2012-102722	First Class	Consumer	106.5
CA-2012-102778	First Class	Consumer	18.176
CA-2012-117828	First Class	Consumer	194.32
CA-2012-130218	First Class	Consumer	59.48
CA-2012-132318	First Class	Consumer	182.91
CA-2012-137974	First Class	Consumer	2298.9
CA-2012-138625	First Class	Consumer	197.72
CA-2012-141327	First Class	Consumer	440.144
CA-2012-149300	First Class	Consumer	32.985
CA-2012-150560	First Class	Consumer	106.62





...using **POWER QUERY** to automate and clean data saves me hundreds of hours of work.





Download the Dataset to get your
hands dirty

[https://foresightbi.com.ng/wp-content/uploads/2020/05/1.-
Badly-Structured-Sales-Data-1.xlsx](https://foresightbi.com.ng/wp-content/uploads/2020/05/1.-Badly-Structured-Sales-Data-1.xlsx)

Don't forget to hit the like button, share, comment and
follow for more.

[linkedin.com/in/olatanrewaju-olanrewaju-5b910480](https://www.linkedin.com/in/olatanrewaju-olanrewaju-5b910480)

 Like  Share  Comment  Save

