DATA PROFILING AND CLEANING

ISSUE	WHERE	PROCESS
Structural Errors		
Size Of Spreadsheet/Excel/Total Number Of Data Fields	ON THE FULL SHEET	27 fields found
Data Format-Wide/Long	ON THE FULL SHEET	wide
Title Check	HEADING SAYS 2013 DATE	Delete it
Check And Remove Extraneous Data	FOUND DATA ABOVE THE SHEET	Delete above rows
Delete Unnecessary Fields Not Relevant To You	FOUND DATA BELOW THE SHEET	Delete it
Remove Extraneous Formatting	FOUND EXTRANEOUS FORMATS	Clear formats
Check Data Types		Check all fields
Count Check		20 rows,7 columns
Range Check		Order ID,Price,Quantity
First Row	REGION ENTERED INCORRECTLY	Do Spell Check
Last Row	CHECKED,IT'S FINE	——————————————-
Spell Check	CHICAGO ENTERED INCORRECTLY	Do Spell Check
Duplicates Check/Unique Check and Removing duplicates	FOUND DUPLICATE ROWS	Data-remove duplicates
Changing Column Names	CHANGE COLUMN NAME CATEGORIES TO CATEGORY	Type correct spelling
Currency Sign rupees to pounds	Pounds	Change into pounds
Dates Format-Alignment	FOUND ALIGNMENT ISSUES ON THE	Format-Alignment

	CUSTOMER NAME COLUMN	
Delete unwanted Columns	FOUND UNWANTED COLUMNS	Yes,Delete it
Replace texts	CHANGE CANDIES NAME TO CANDY UNDER CATOGORIES COLUMN NAME	Find-Replace
Extra space in between the customer names in the in Customer Name column	Trim-credit card	Use trim function
Missing Values		
● Count Total Of Blank Cells		Go to special-Blank
● Partial Or Incomplete Values in Customer Names	Yes, Daniel	Check with source provider
● Highlight Missing Spaces	Yes, found it	Check with source provider to get the data
Data Outside Range		
● Quantity Values Less Than 18 And Greater Than 100	Yes, found. Quantity 19000	Check with source provider
● Dates On Transactions Earlier Than The Time The Business Started	Yes 2013 year	Check with source provider
● Codes, Like Postal Codes Or Zip Codes, That Are Outside The Range Of Defined Values	Yes issues	Input with correct postcodes OR Check with source provider
Malformed Data		
● Names that Are Truncated	daniel	Check with source provider
● Check For A Middle Initial	Yes, Found.Customer name.Brian Tom Taylor	Solved with Text to columns

● Dates Formatted With Invalid Characters/Strip Out Undesirable Characters	yes*******8 on City column	Find- Replace option
● Standardising case	Yes.in customer name	Use String Functions -PROPER Function
Find Length Of Words	Check order iD, Quantity, Shipping Price Quantity shows 19,000 Under Products-Sauces shows -400	Check with source provider Check with source provider
Split Date First And Time	Yes-name date and time	Use Text to columns option
Business Rules		
● An Order Total Is Greater Than A Customer's Credit Limit/Purchasing limit	Sauces -400	Check with source provider
● Values That Are Larger Or Smaller Than A Cutoff Value/Finding Outliers	Quantity shows 19,000 Under Products-Sauces shows -400	Check with source provider
● Values That Are Negative Or Positive	SOPHIA -84 IN QUANTITY	Check with source provider
● Date Range	2013 DATE ENTRY FOUND	Check with source provider
● Date Format	LONG DATE FORMAT FOUND	Format date into short form
● Errors In The Data Or In Calculations That Drive Values In The Data	yes	After getting the correct data from the source provider,do the revenue calculation of each.

		OR Delete the column if unnecessary.
● Removes sum	FOUND SUM ₹24,671.67	Delete it
● Currency Operations	NEED TO CHANGE RUPEES TO POUNDS IN REVENUE COLUMN	Home-Select appropriate currency
● The Date An Order Is Shipped Is Earlier Than The Date It Was Ordered	A DATE IS FOUND BELONGS TO 1381 ORDER ID COLUMN IT'S A 2013 YEAR DATE	Check with source provider
Inconsistent Values		
● Order Date May Have An Order Id	yes	Do the appropriate removing procedure OR If needed check with source provider
● Order Items Are Associated With An Order	Need to check	Check with source provider
● An Order Item Has An Order Id That Does Not Exist In The Order Table	FOUND CORN FLOUR UNDER PRODUCT NAME COLUMN	Check with source provider