

# Team Genesis

Round\_2: Envision

## Data Cleaning and Solution Planning

### Step-1: Data Cleaning

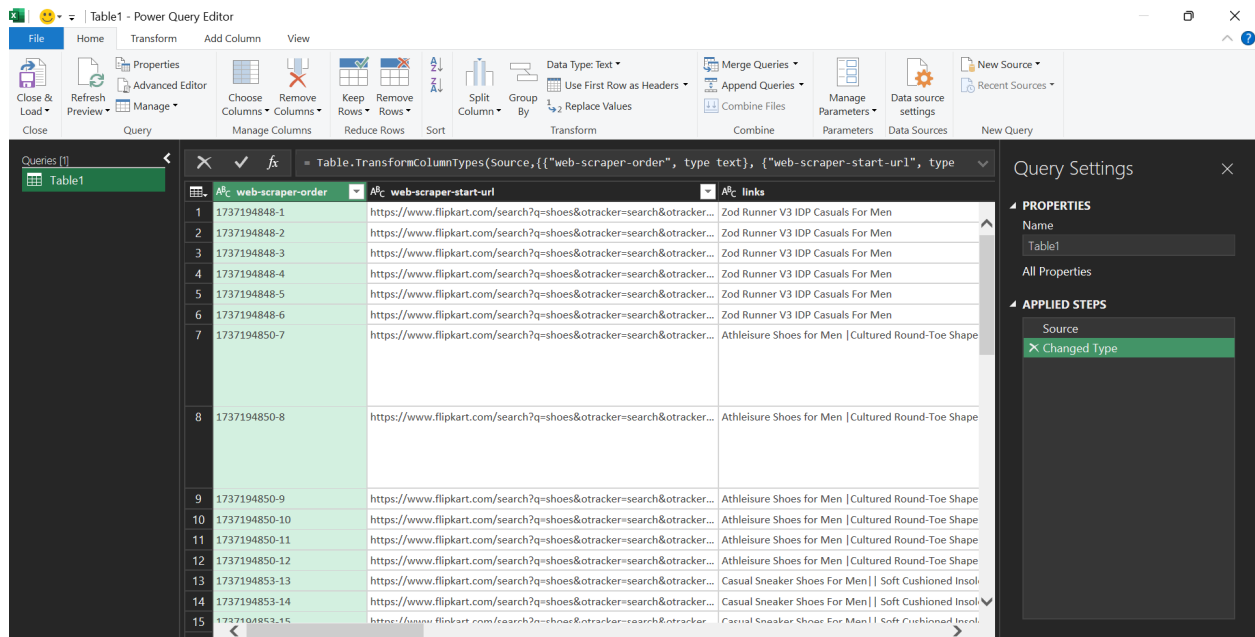
#### 1.1: Date Formatting:

The screenshot shows an Excel spreadsheet with the following data:

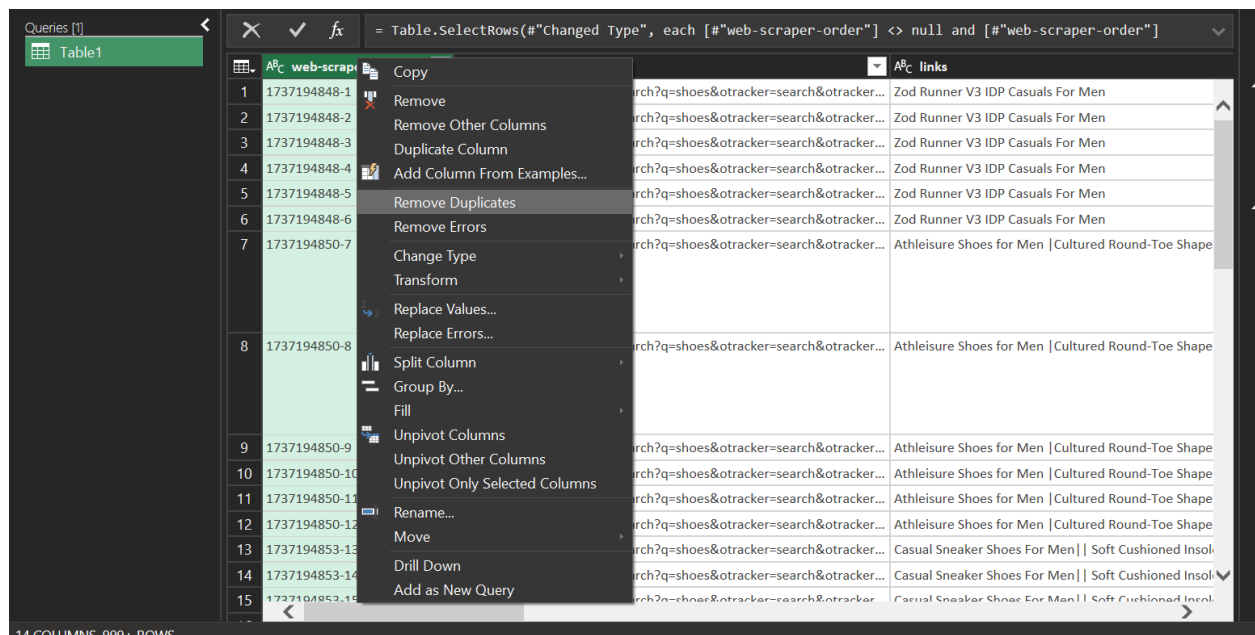
	links	links-href	name	actual_pri	discount	discount	product_rating	user_name
2	Boult Z40	https://wn	Boult Z40	₹ 4,999	₹ 899	82% off	4.1	Santhosh Sandy
3	Boult Z40	https://wn	Boult Z40	₹ 4,999	₹ 899	82% off	4.1	Apurv Gholap
4	Boult Z40	https://wn	Boult Z40	₹ 4,999	₹ 899	82% off	4.1	Apurv Gholap
5	Boult Z40	https://wn	Boult Z40	₹ 4,999	₹ 899	82% off	4.1	Apurv Gholap
6	Boult Z40	https://wn	Boult Z40	₹ 4,999	₹ 899	82% off	4.1	Flipkart Customer
7	Boult Z40	https://wn	Boult Z40	₹ 4,999	₹ 899	82% off	4.1	Flipkart Customer
8	Boult Z40	https://wn	Boult Z40	₹ 4,999	₹ 899	82% off	4.1	Flipkart Customer
9	Boult Z40	https://wn	Boult Z40	₹ 4,999	₹ 899	82% off	4.1	MD Sohel Dandin
10	Boult Z40	https://wn	Boult Z40	₹ 4,999	₹ 899	82% off	4.1	MD Sohel Dandin
11	Boult Z40	https://wn	Boult Z40	₹ 4,999	₹ 899	82% off	4.1	MD Sohel Dandin
12	Boult Z40	https://wn	Boult Z40	₹ 4,999	₹ 899	82% off	4.1	Arivazhagan Aris
13	Boult Z40	https://wn	Boult Z40	₹ 4,999	₹ 899	82% off	4.1	Arivazhagan Aris
14	Boult Z40	https://wn	Boult Z40	₹ 4,999	₹ 899	82% off	4.1	Arivazhagan Aris
15	Boult Z40	https://wn	Boult Z40	₹ 4,999	₹ 899	82% off	4.1	Flipkart Customer
16	Boult Z40	https://wn	Boult Z40	₹ 4,999	₹ 899	82% off	4.1	Flipkart Customer
17	Boult Z40	https://wn	Boult Z40	₹ 4,999	₹ 899	82% off	4.1	Flipkart Customer
18	Boult Z40	https://wn	Boult Z40	₹ 4,999	₹ 899	82% off	4.1	Minakshi Rout
19	Boult Z40	https://wn	Boult Z40	₹ 4,999	₹ 899	82% off	4.1	Minakshi Rout
20	Boult Z40	https://wn	Boult Z40	₹ 4,999	₹ 899	82% off	4.1	Minakshi Rout
21	Boult Z40	https://wn	Boult Z40	₹ 4,999	₹ 899	82% off	4.1	Leonard Aomeshu
22	Boult Z40	https://wn	Boult Z40	₹ 4,999	₹ 899	82% off	4.1	Leonard Aomeshu
23	Boult Z40	https://wn	Boult Z40	₹ 4,999	₹ 899	82% off	4.1	Leonard Aomeshu

The 'Format Cells' dialog box is open, showing the 'Date' category selected. The 'Type' list shows various date formats, with '14-03-2012' selected. The 'Locale (location)' is set to 'English (India)' and the 'Calendar type' is set to 'Gregorian'.

Used Power Query Model to Clean and Process data:



Removing duplicates:



**Data Preprocessing and Visualization:**

<https://github.com/Meenalsh72/Envision/blob/main/envision.ipynb>

data processing and visualization are mentioned in the ipynb notebook above

## Solution Overview:

### Customer Segmentation for Marketing Analysis

This project aims to segment customers based on purchasing behavior to enhance marketing strategies. By identifying key customer groups, businesses can personalize campaigns, improve retention, and maximize ROI.

---

### Why RFM Analysis?

RFM (Recency, Frequency, Monetary) analysis evaluates customers on:

- **Recency:** How recently they purchased.
- **Frequency:** How often they purchase.
- **Monetary:** How much they spend.

RFM provides a clear framework for identifying high-value customers, loyal buyers, and churn risks, forming the foundation for actionable segmentation.

---

### Why DBSCAN?

DBSCAN is ideal for clustering RFM data because:

- It handles outliers (e.g., sporadic buyers).
  - Captures irregularly shaped clusters that match real-world customer behavior.
  - Requires no predefined number of clusters, offering flexibility.
-