

---

# My Final Project: Association Rule Mining with R

**Student ID:** 79451

**Name:** Michael tefera

---

## Q1: How many unique items are there?

There are **29 unique ingredients** across all transactions in the dataset.

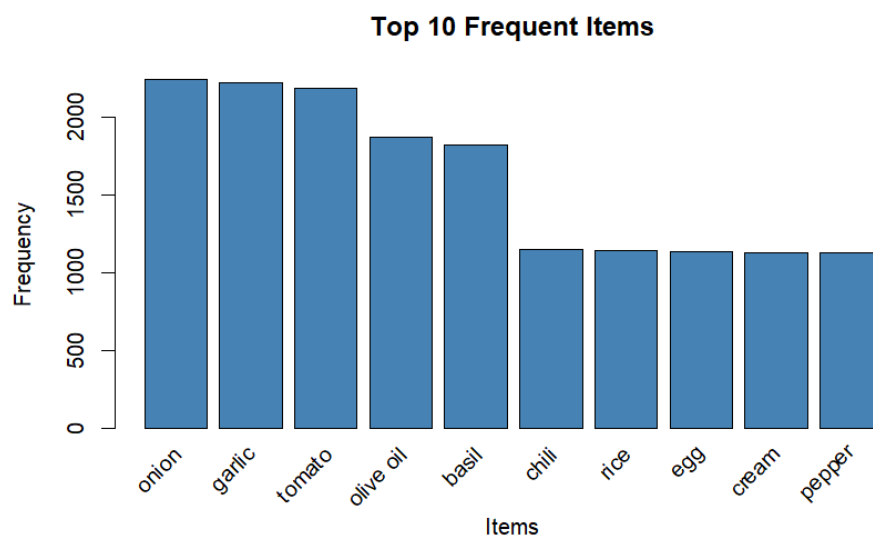
---

## 2: What are the top 3 most frequent items?

The top 3 most frequent ingredients :

1. **Onion** – 2,243 occurrences
2. **Garlic** – 2,223 occurrences
3. **Tomato** – 2,185 occurrences

*Plot: Top 10 Frequent Items*



---

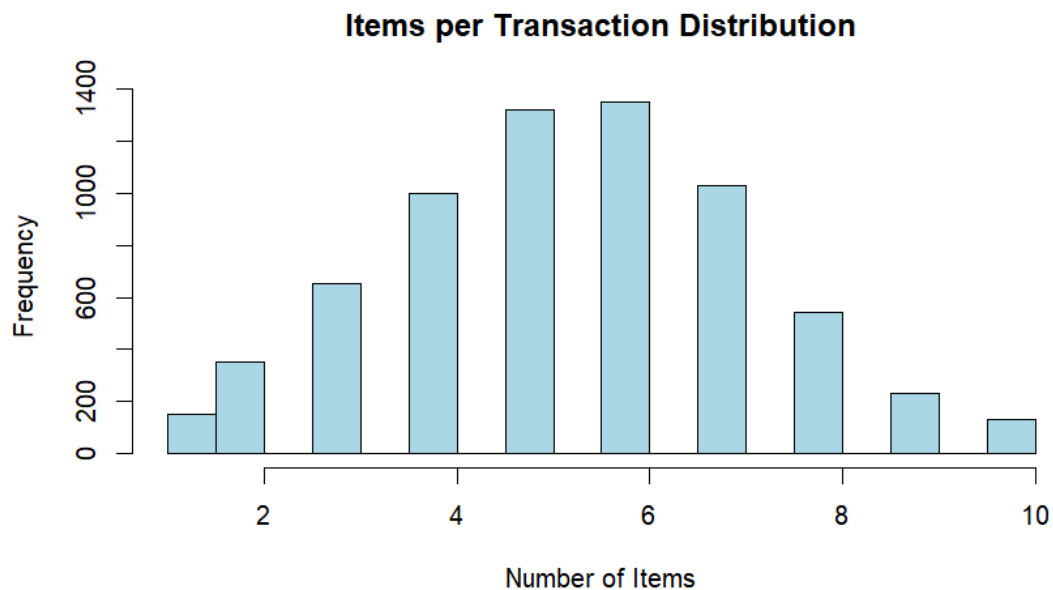
### 3: How many transactions are there? What is the distribution of the number of itemsets per transaction?

Answer:

- Total number of transactions: **6,750**
- Average items per transaction: **5.39**

The distribution of item counts per transaction is shown below.

*Plot: Items per Transaction Histogram*



---

### 4: What are the top 3 frequent 1-itemsets?

Item	Support	Count
Onion	0.332	2,243
Garlic	0.329	2,223
Tomato	0.324	2,185

---

### 5: What are the top 3 frequent 2-itemsets?

Itemset	Support	Count
Garlic & Tomato	0.215	1,450
Garlic & Onion	0.213	1,441
Onion & Tomato	0.211	1,424

---

### 6: What are the top 3 frequent 3-itemsets?

Itemset	Support	Count
Garlic, Onion, Tomato	0.191	1,292
Garlic, Olive Oil, Tomato	0.028	187
Garlic, Salt, Tomato	0.027	185

---

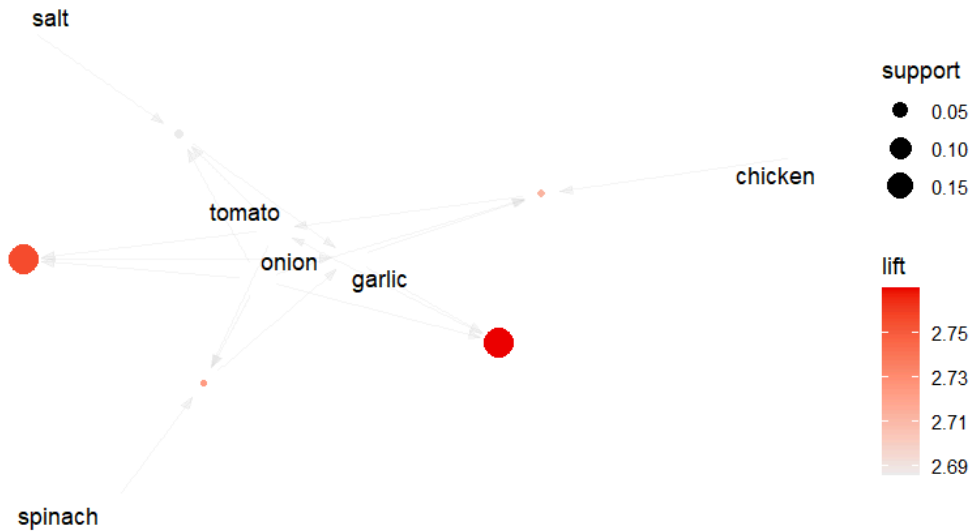
### 7: Generate association rules with support = 1% and confidence = 50%. Show top 5 rules.

Rule	Support	Confidence	Lift
Garlic & Onion → Tomato	0.191	0.897	2.77
Onion & Tomato → Garlic	0.191	0.907	2.75
Onion, Spinach & Tomato → Garlic	0.021	0.897	2.72
Chicken, Garlic & Onion → Tomato	0.021	0.878	2.71
Onion, Salt & Tomato → Garlic	0.023	0.884	2.69

---

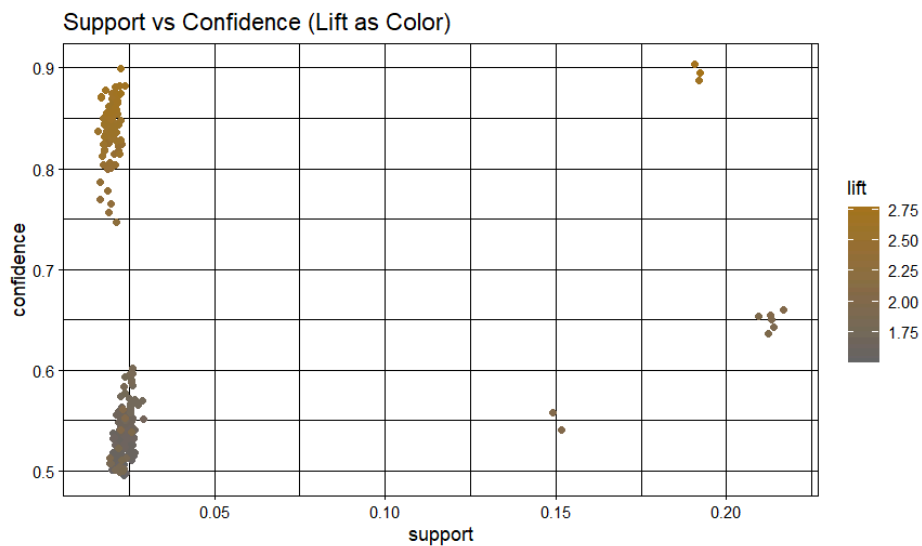
## Q8: Plot the rules as a graph to visualize relationships (LHS → RHS)

The graph below shows the **top 5 association rules**, with arrows from antecedents (LHS) to consequents (RHS). Node size indicates item frequency, and edge thickness reflects rule strength.



## Q9: Plot a scatterplot of rules with support and confidence on the axes, and lift as the shading color

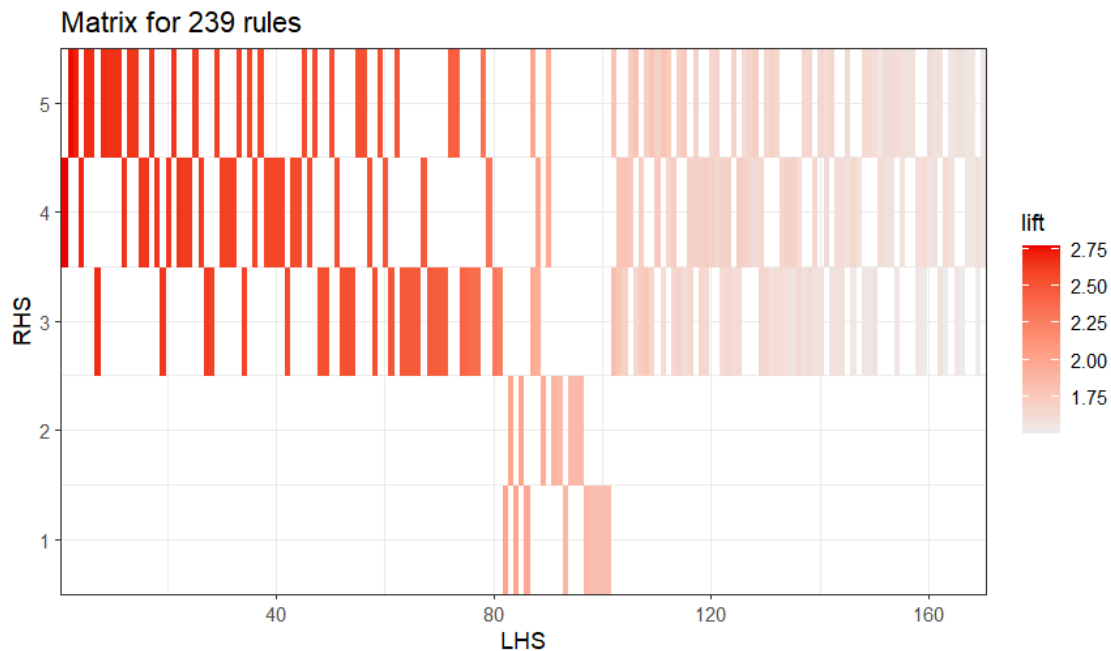
This scatterplot maps support (x-axis) vs confidence (y-axis), using **lift as the color shading**. Darker points represent stronger rules.



---

## 10: Plot the rules as a matrix plot, using lift and confidence as measures

This matrix visualizes rule **confidence** (horizontal axis) and **lift** (color intensity), helping to highlight strong rules.



---

## Q11: Generate new rules with support = 1% and confidence = 40%

Answer:

- Number of rules with 50% confidence: **239**
- Number of rules with 40% confidence: **295**

---

## Q12: Explain what happens when minimum support is increased

Increasing the minimum support threshold **reduces** the number of generated rules. This is

because itemsets that occur less frequently are **filtered out**.

While this improves **rule reliability**, it can also cause us to **miss rare but potentially interesting patterns**.

---