

Fashion E-Commerce in the USA



Data Sets

Consumer Behavior and Shopping Habits Dataset (E-Commerce Transaction Trends)

- 3900 unique customer IDs over variables such as: age, gender, item purchased, category, location, season, buy frequency...)
- Source: **Kaggle**

Income by US State Dataset (Per Capita Personal Income)

- Comprising all 50 states
- Source: **Bureau of Economic Analysis**

Business Goal: Increasing sales

-> Insights into consumers' preferences, tendencies, and patterns during their shopping experiences.

H1: Customer buy more fashion items in the summer.

H2: Different seasons have different types of top seller categories (zB: Outwear -> Fall)

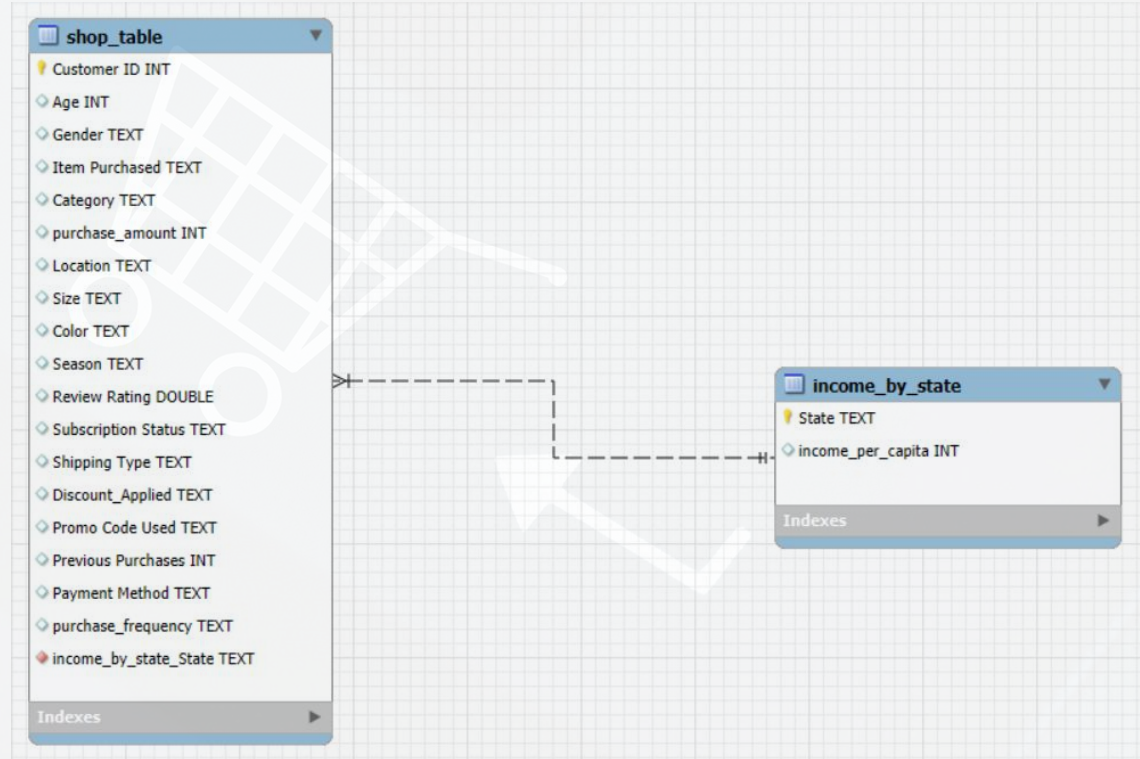
H3: Female customers purchase more Clothing items than male customers

H4: Younger customers purchase more fashion items than older customers

H5: Income per capita correlates positively with purchase frequency & average amount spent per item

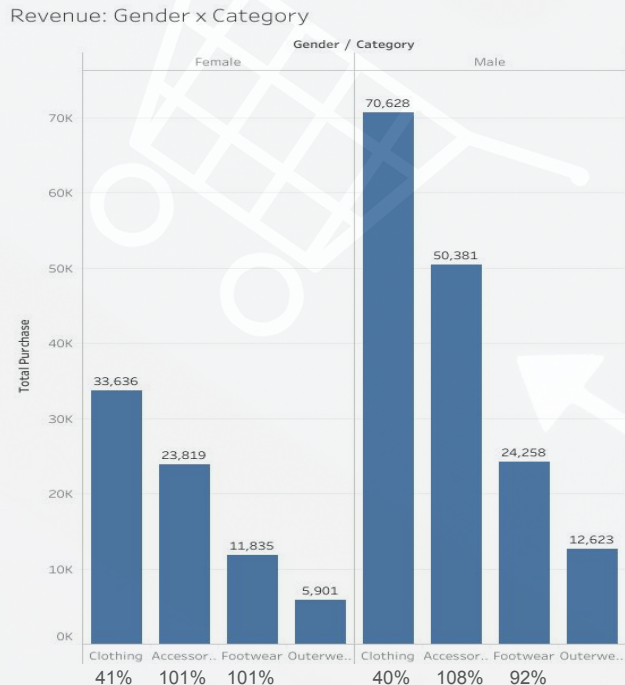
Entity-Relational-Model & Key Challenges

- Connecting two existing tables through keys proved to be a challenge using EER Diagrams
- We solved this by enforcing the join within the SQL Script (ON)



Major Obstacles

- Not much context to shopping data set source, likely synthetic data



SQL Queries Insights

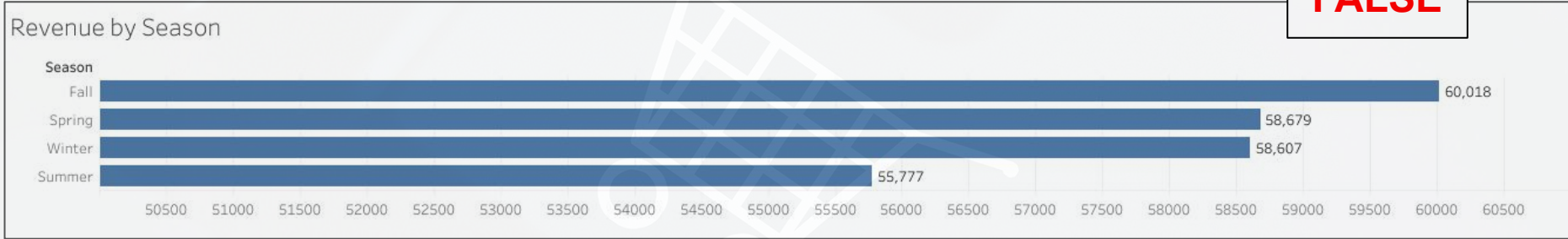
```
SELECT
  CASE
    WHEN purchase_frequency = "Bi-Weekly" OR "Weekly" THEN "Weekly"
    WHEN purchase_frequency = "Fortnightly" OR "Monthly" OR "Every 3 Months" THEN "2 Weeks - 3 Months"
    ELSE "More than 3 months"
  END AS buying_frequency,
  COUNT(*) AS count,
  SUM(purchase_amount) AS total_purchase,
  SUM(purchase_amount)/COUNT(*) AS average_purchase_amount
FROM shop_table
GROUP BY buying_frequency
ORDER BY buying_frequency;
```

```
SELECT
  CASE
    WHEN age < 18 THEN 'Under 18'
    WHEN age BETWEEN 18 AND 29 THEN '18-29'
    WHEN age BETWEEN 30 AND 49 THEN '30-49'
    ELSE '50 and above'
  END AS age_group,
  category,
  SUM(purchase_amount) AS total_purchase
FROM shop_table
GROUP BY age_group, category
ORDER BY age_group, category;
```

```
SELECT state, income_per_capita, COUNT(purchase_frequency) AS purchase_frequency, SUM(purchase_amount)/COUNT(purchase_frequency) AS average_purchase_amount
FROM income_by_state
INNER JOIN shop_table
ON income_by_state.state = shop_table.location
GROUP BY state, income_per_capita
ORDER BY income_per_capita DESC;
```

H1: Customer buy more fashion items in the summer.

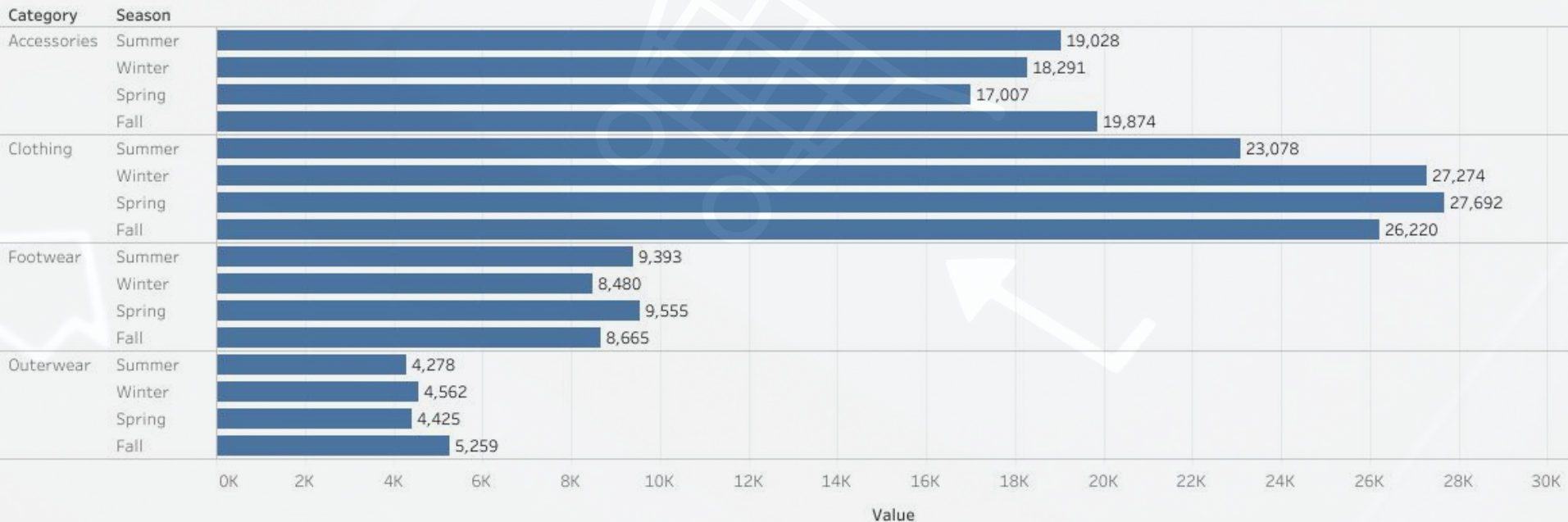
FALSE



H2: Different seasons have different types of top seller categories

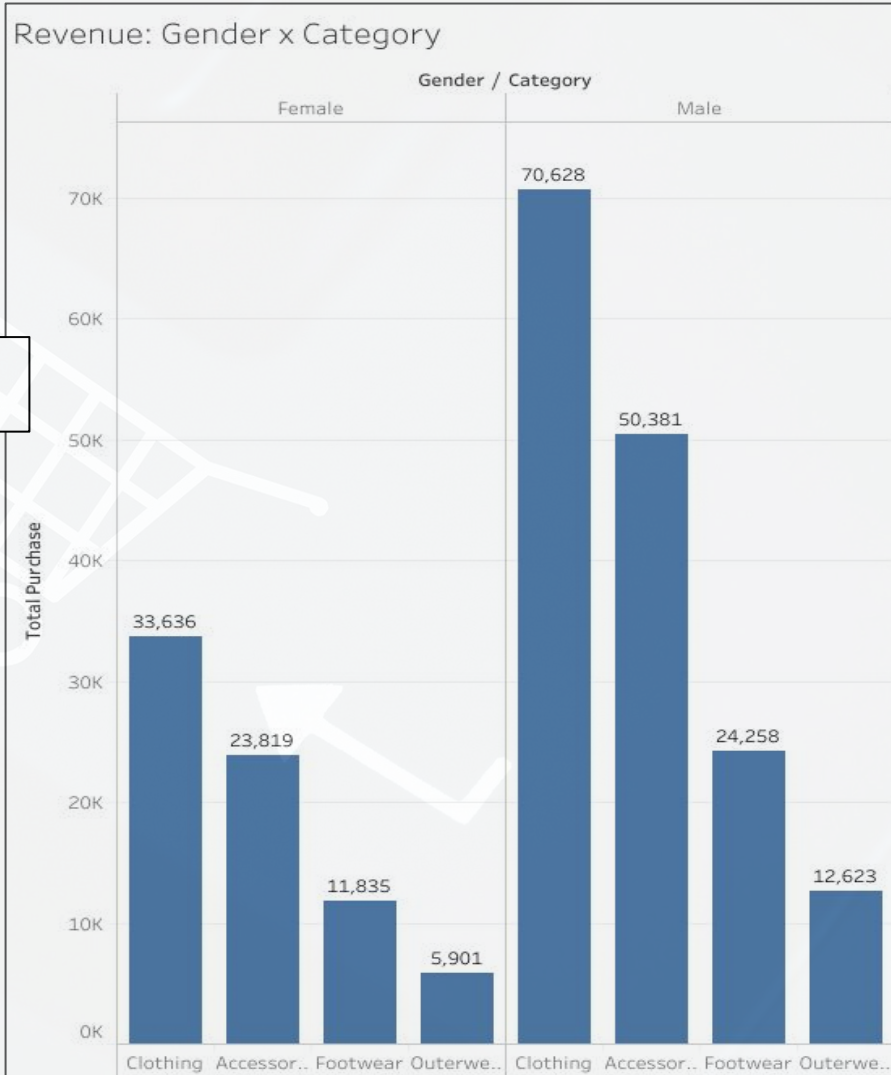
TRUE

Revenue, by Season x Category



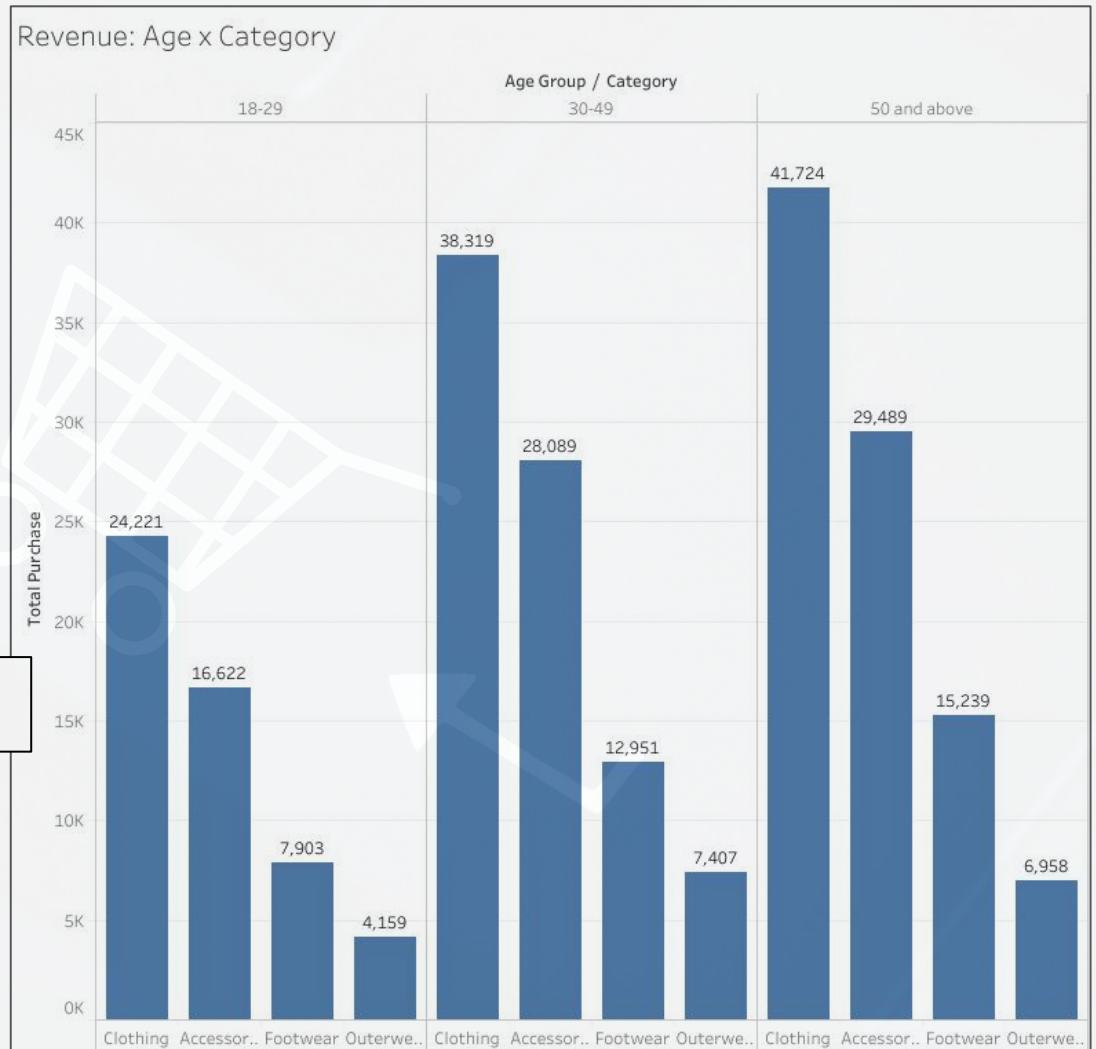
H3: Female customers purchase more Clothing items than male customers

FALSE



H4: Younger customers purchase more fashion items than older customers

FALSE

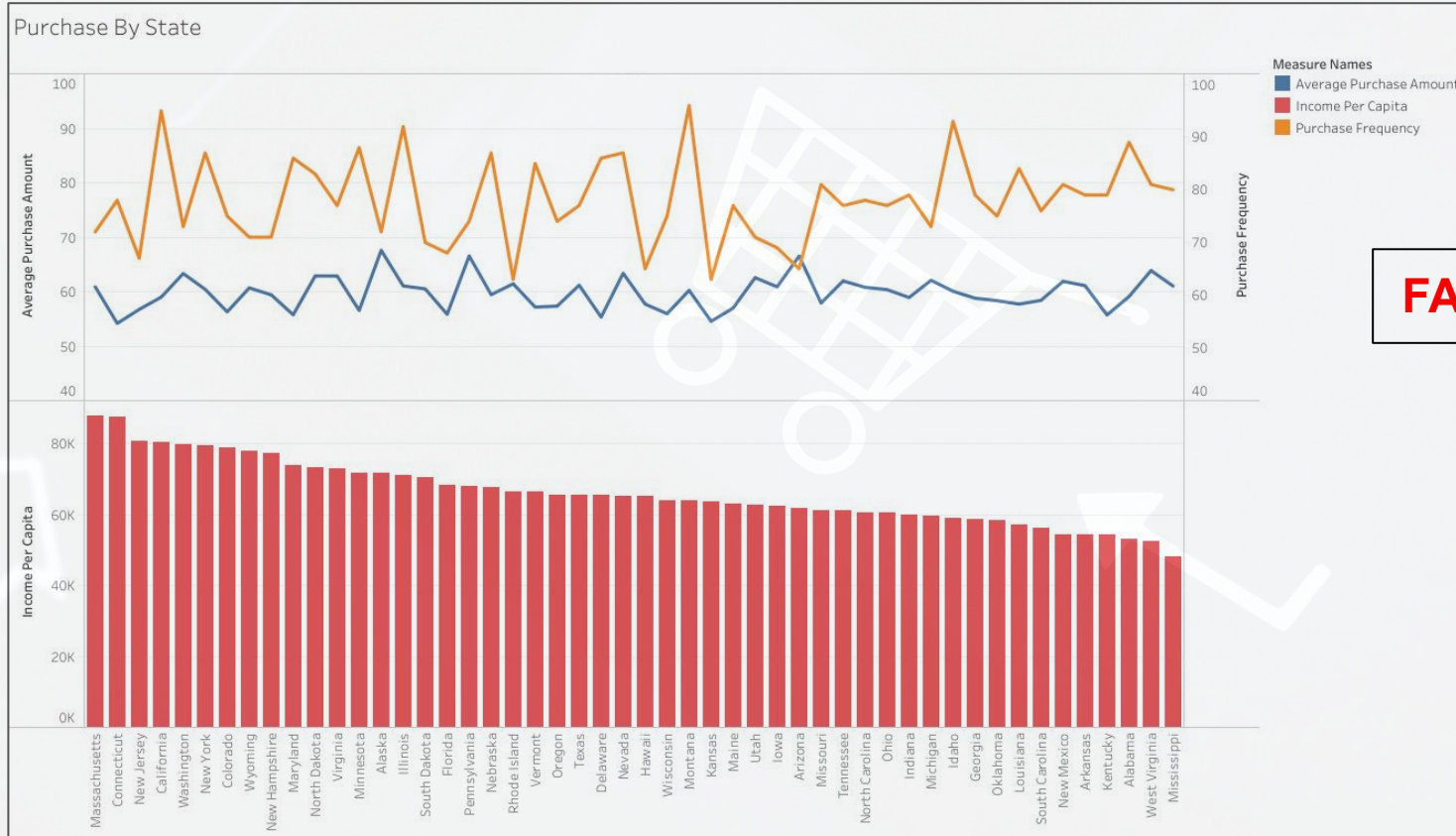


Insights

Purchase Frequency: Count x Average Purchase



H5: Income per capita correlates positively with purchase frequency & average amount spent per item



FALSE

Conclusion

Business Goal: Increasing sales

-> Insights into consumers' preferences, tendencies, and patterns during their shopping experiences.

Hypothesis 1: Customer buy more fashion items in the summer.

-> FALSE

Hypothesis 2: Different seasons have different types of top seller categories

-> TRUE

Hypothesis 3: Female customers purchase more Clothing items than male customers

-> FALSE

Hypothesis 4: Younger customers purchase more fashion items than older customers

-> FALSE

Hypothesis 5: Income per capita correlates positively with purchase frequency & average amount spent per item

-> FALSE