

# Project 1 – Ecommerce Database

BY MELISSA NIELSEN



# Getting to Know Data

- Determined that all info I needed was in all\_sessions
- Only 81 rows had 'totaltransactionrevenue' not null, so focused on these as the *sales*
- Returned distinct count for all columns and stored this data in excel – this was very helpful for understanding the data
- Determined that other tables did not contain data which would help me look at sales by city, country, etc



# Scope Determination

Investigate behaviour and demographics (e.g. location) of site visitors who had viewed and purchased products

# Key Assumptions

Only rows with a transaction revenue not equal to zero had purchased products

# Data Cleaning

- Found over 3 million rows of duplicate data in analytics! But decided that this dataset was not helpful to me
- Updated currency code based on country
- Divided all financial figures by 1,000,000
- Fixed location info for visitors who had made sales (81 rows only)
- Updated product categories
- Some cities missing and indicated 'not available in demo dataset' these were not included in city analyses



# Data Cleanup – Product Categories

- Updated product categories to list on right
- Mostly used 'LIKE' operator with '%value%'
- Focused on rows where totaltransactionrevnue not null

Accessories

Apparel

Bags

Drinkware

Electronics

Lifestyle

Nest

Office

Waze



# DATA CLEANUP

SOMETIMES THE  
PRODUCT CATEGORY IS  
AMBIGUOUS!

ARE WAZE DRESS SOCKS  
APPAREL OR WAZE?

WHO BUYS WAZE DRESS  
SOCKS ANYWAY?!?





# Key Data Cleanup Takeaway

Product quantities were unusable

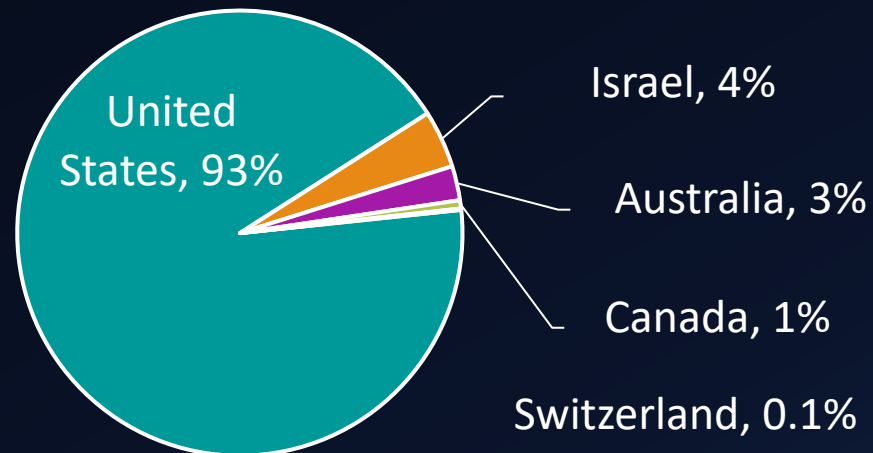
- Only 18 transactions included product quantities and they were not correlated to transaction revenue by their quantities and product prices

# Results (starting\_with\_questions)

- Q1 - Which cities and countries have the highest level of transaction revenues on the site?
  - United States had most transaction revenue by far - \$13K out of \$14K total revenue was from the United States
  - Atlanta had the most transaction revenue (\$850)
- Q2 –What is the average number of products ordered from visitors in each city and country?
  - Data was **insufficient** to answer this question (most product quantities missing)
- Q3 – Is there any pattern in the types (product categories) of products ordered from visitors in each city and country?
  - United States was the only country with more than one transaction
  - The most commonly purchased categories in the United States were
    - Apparel (27 out of 77)
    - Nest (27 out of 77)

## Results (starting\_with\_questions)

- Q4 - What is the top-selling product from each city/country? Can we find any pattern worthy of noting in the products sold?
  - The top-selling product in the United States is 'Nest® Learning Thermostat 3rd Gen-USA - Stainless Steel' with seven transactions. All other countries have only one transaction
  - There were no instances of an item being purchased twice in any one city.
- Question 5: Can we summarize the impact of revenue generated from each city/country?



# Additional Investigations

- How many visitors viewed or purchased more than one product?
  - Did these visitors come back on multiple days?
- How much time did visitors spend on the site?



## Results (additional questions)

I chose to answer the following questions:

Q1 - Find the number of visitors who viewed more than one product type.

Q2 - Find the number of visitors who purchased more than one product type.


Q3 - How much time did visitors spend on the site who DID purchase items vs those who did not?

Q4 - How many users spent less than 1 minute seconds on the site?

Q5 - For visitors who viewed more than 1 product, did they come back to the site on a different day?

# Visitors who viewed and purchased more than one item

Interaction Type	One Item Only	More than One Item
Viewed	13,429	805
Purchased	79	1



# For visitors who viewed more than 1 product, did they come back to the site on a different day?

- Majority of visitors who viewed more than one product visited the site on one day only
  - 805 viewed more than once
    - 566 of these viewed site on one day only



## Time on Site

Interaction Type	Average Time on Site
Viewed Only	3.7 minutes
Purchased	9.4 minutes

- 3,729 visitors spent less than one minute on the site.
- *Note: I assumed that time on site was measured in seconds – these average site visit times seem surprisingly long!*

# QA Process

- Key numbers that I used in the QA process were:

Description	Count
Count of all rows	15,134
Count of unique fullvisitorid	14,223
Count of unique visitid	14,556
Count of sales	81

- Process typically involved wrapping my code in a subquery, summing my results column(s) and ensuring that the number of results added up to the appropriate count in the table above.

# Future Work



- I focused on cleaning the data for the rows that had an associated revenue (ie. rows where there were sales)
  - *Especially* for the location data and the product categories, since these rows were more difficult to clean
- Future investigations:
  - Understand product quantities – is this data available elsewhere?
  - What is the difference between visits that DID turn into a sale vs those that DID NOT?
    - Would require a more thorough cleaning of location and product category data for *all* rows
  - How did the visitors get to the site? (investigate channelgrouping column)
  - Were product ratings (in the products table) associated with better selling products?