Transforming & Analyzing Data with SQL

A Preliminary Exploratory Data Analysis

Nico San Pedro

Agenda

- Loading the Data and Cleaning the Data
- QA Process
- Preliminary Insights
- Recommendations

Loading the Data and Cleaning the Data

Loading the Data

1. Used pgAdmin's GUI to load the datasets
2. Create the same number of columns on pgAdmin as Character Varying
3. Name the table as raw tables

Cleaning the Data

1. Correcting the Data Type
    a. Creating new tables with the same number of columns as VARCHAR
    b. Inserting the raw data from the raw tables to the new tables
    c. Alter the data type of each column

Cleaning the Data

2. Correcting the Values
    a. Divided the prices by 1,000,000.
    b. Extracted the keyword for productcategory

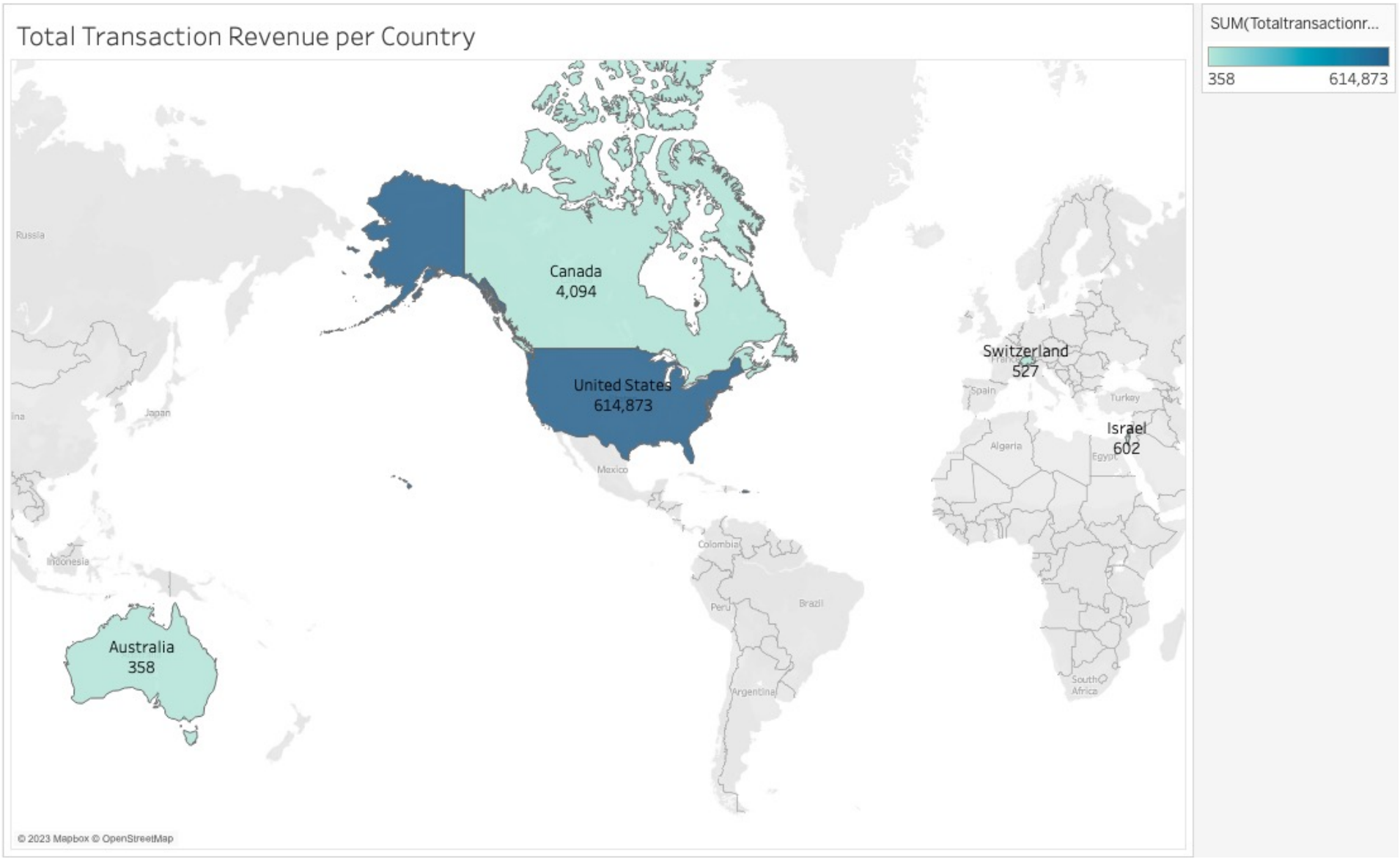
QA Process

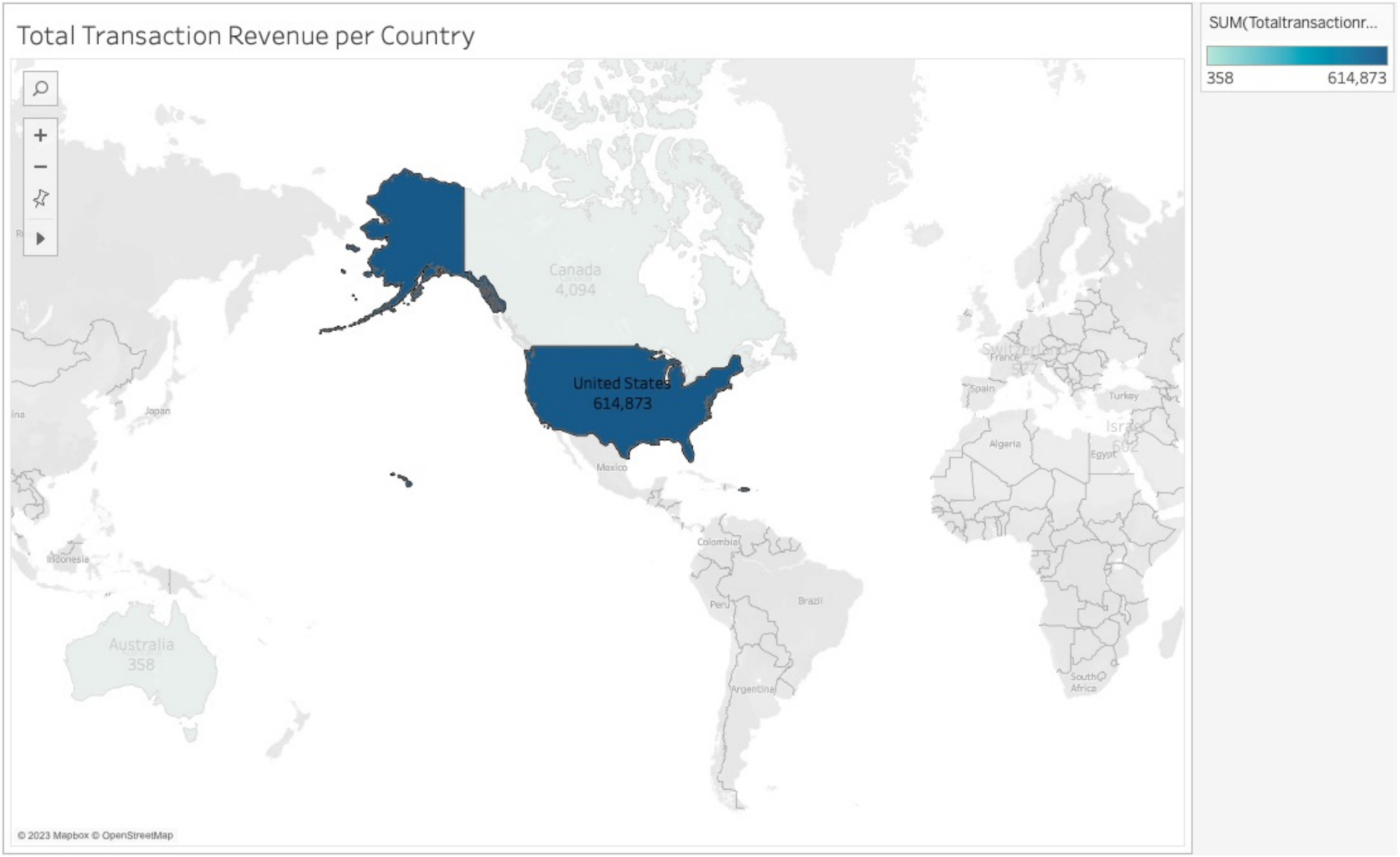
QA Process

1. Created a separate table for the raw datasets.
2. Created a specific query file for cleaning each table.
3. Created PK and added some FK.

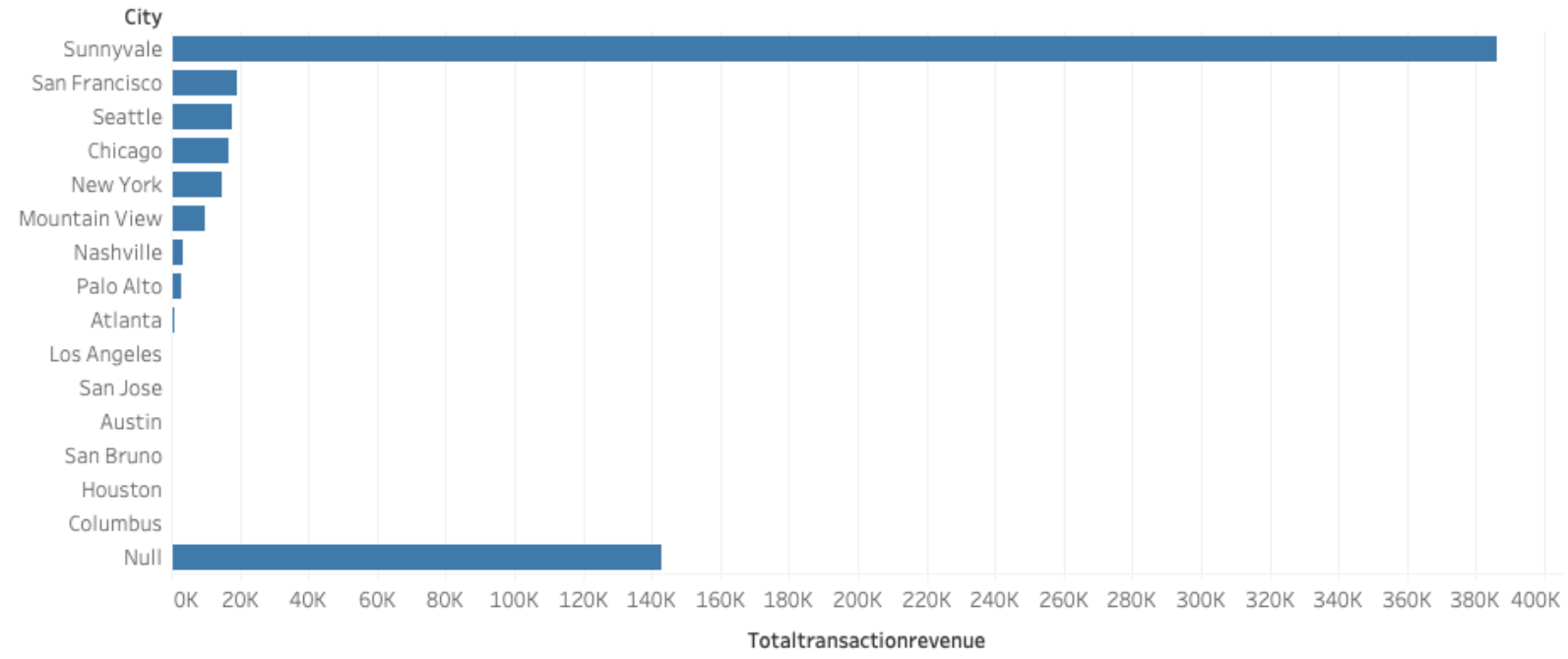
Project Questions

1. Which cities and countries have the highest level of transaction revenues on the site?

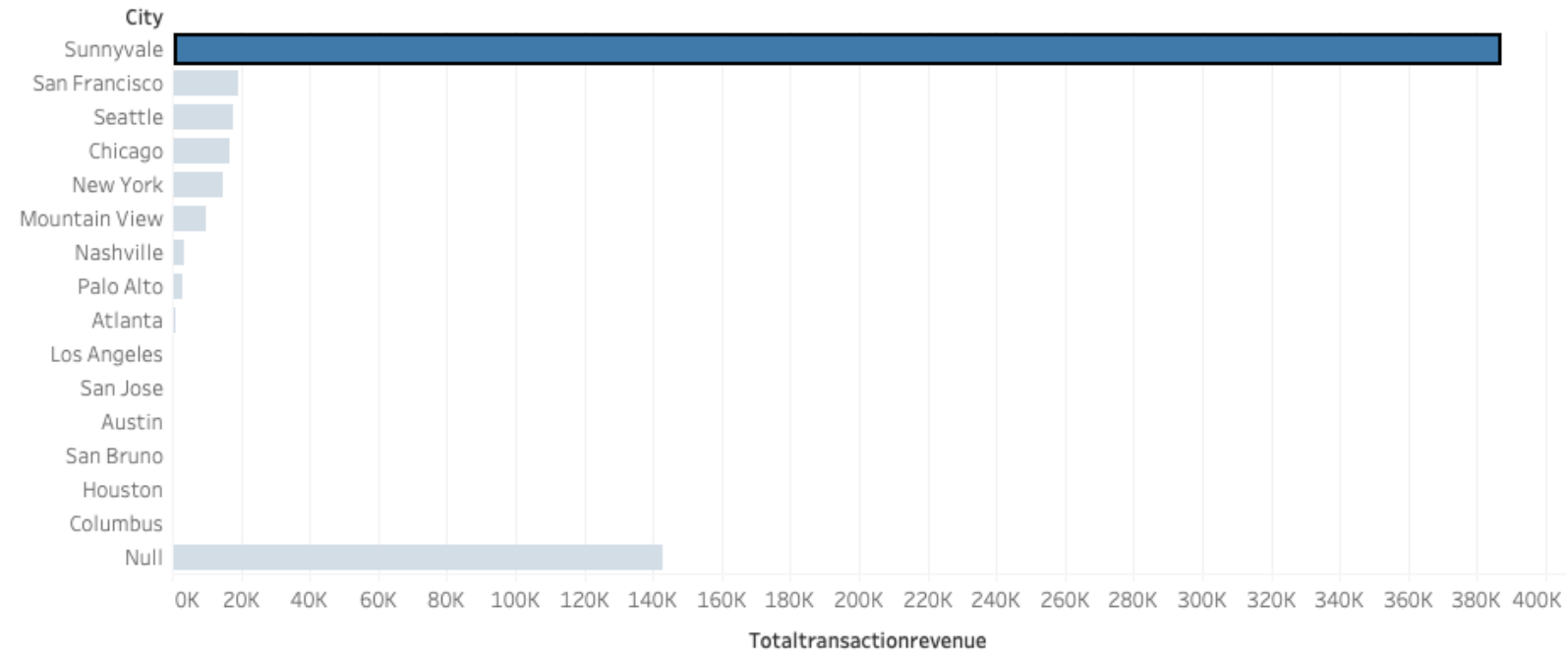

Total Transaction Revenue per Country
SUM(Totaltransactionr...
358
614,873
Canada
4,094
United States
614,873
Switzerland
527
Israel
602
Australia
358
Russia
Japan
Mexico
Colombia
Peru
Brazil
Argentina
Spain
Turkey
Algeria
Egypt
South Africa
Indonesia
© 2023 Mapbox © OpenStreetMap


Total Transaction Revenue per Country
SUM(Totaltransactionr...
358
614,873
Canada
4,094
United States
614,873
Japan
Mexico
Colombia
Peru
Brazil
Argentina
Australia
358
Spain
France
Turkey
Algeria
Egypt
South Africa
© 2023 Mapbox © OpenStreetMap

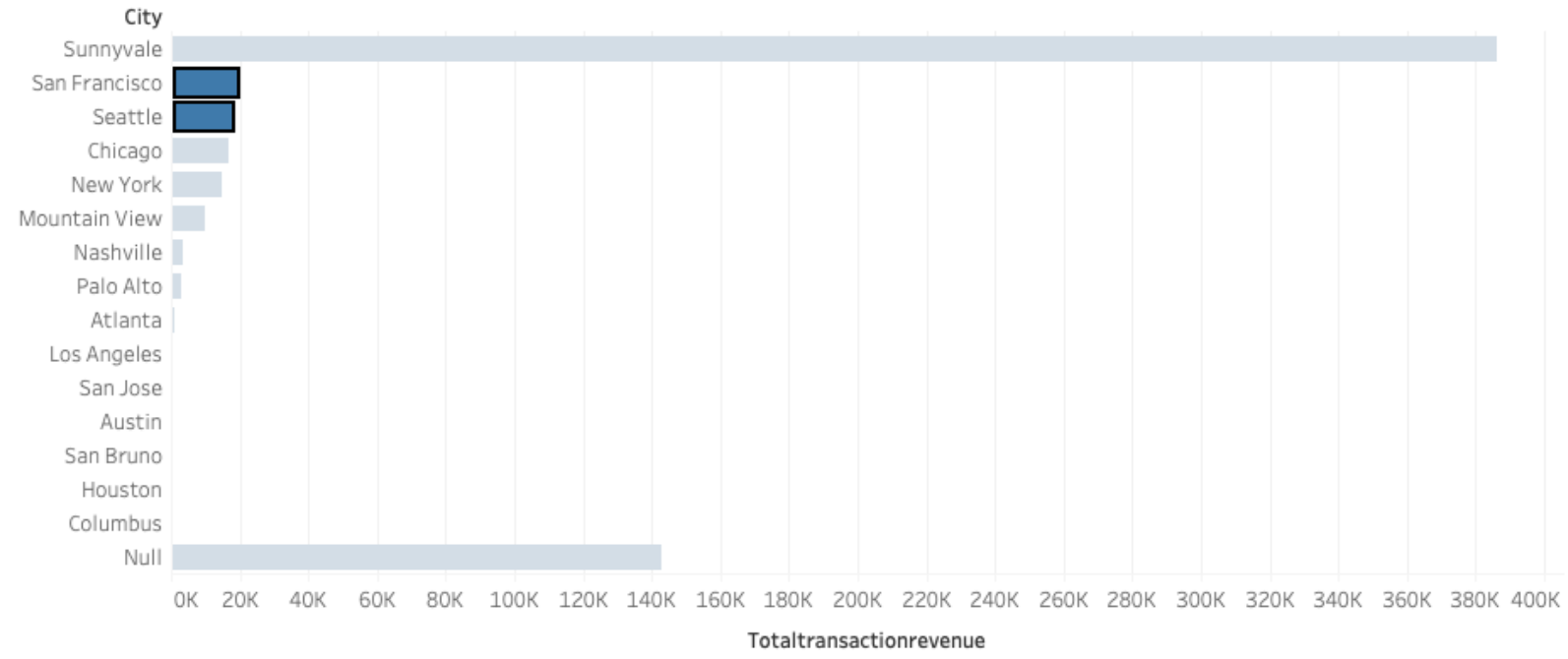
Total Transaction Revenue per City in the USA



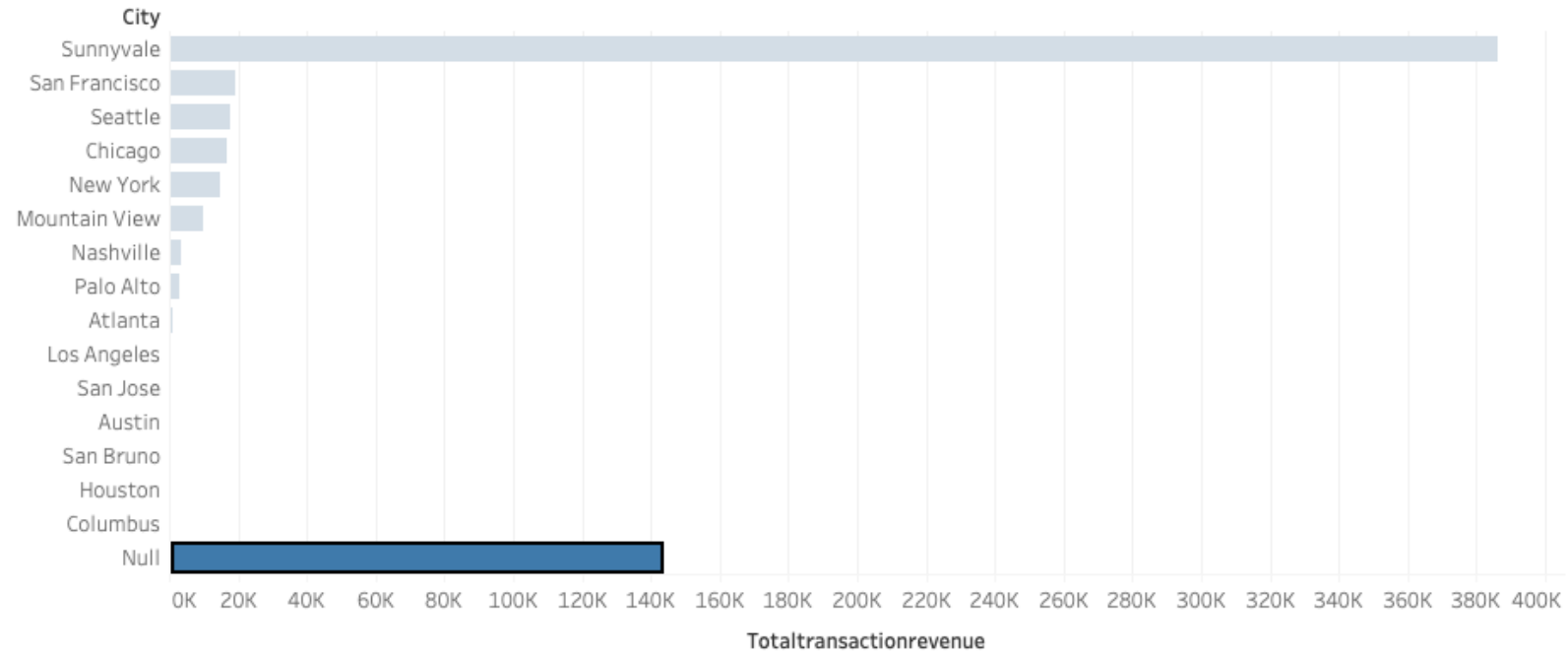
Total Transaction Revenue per City in the USA

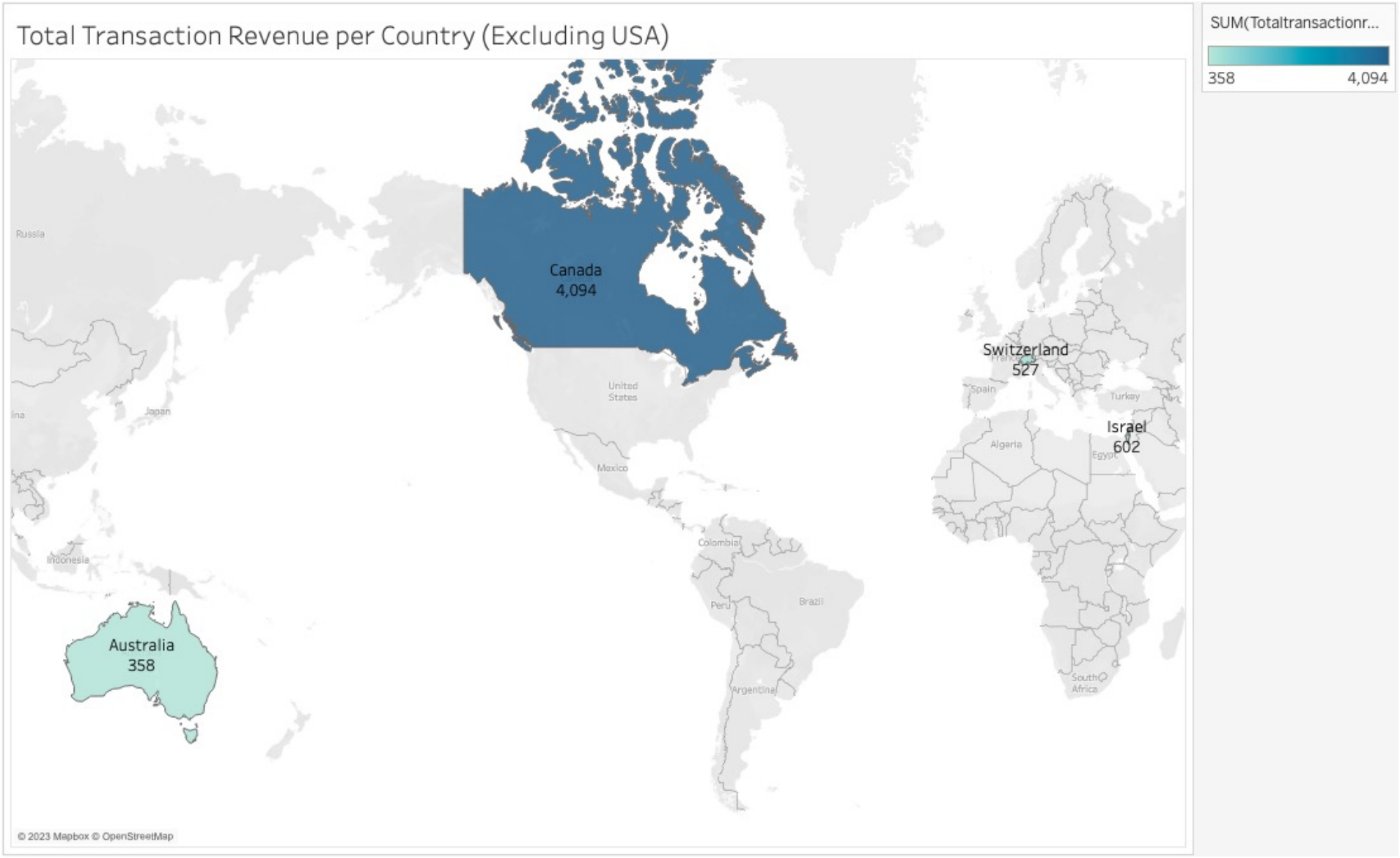


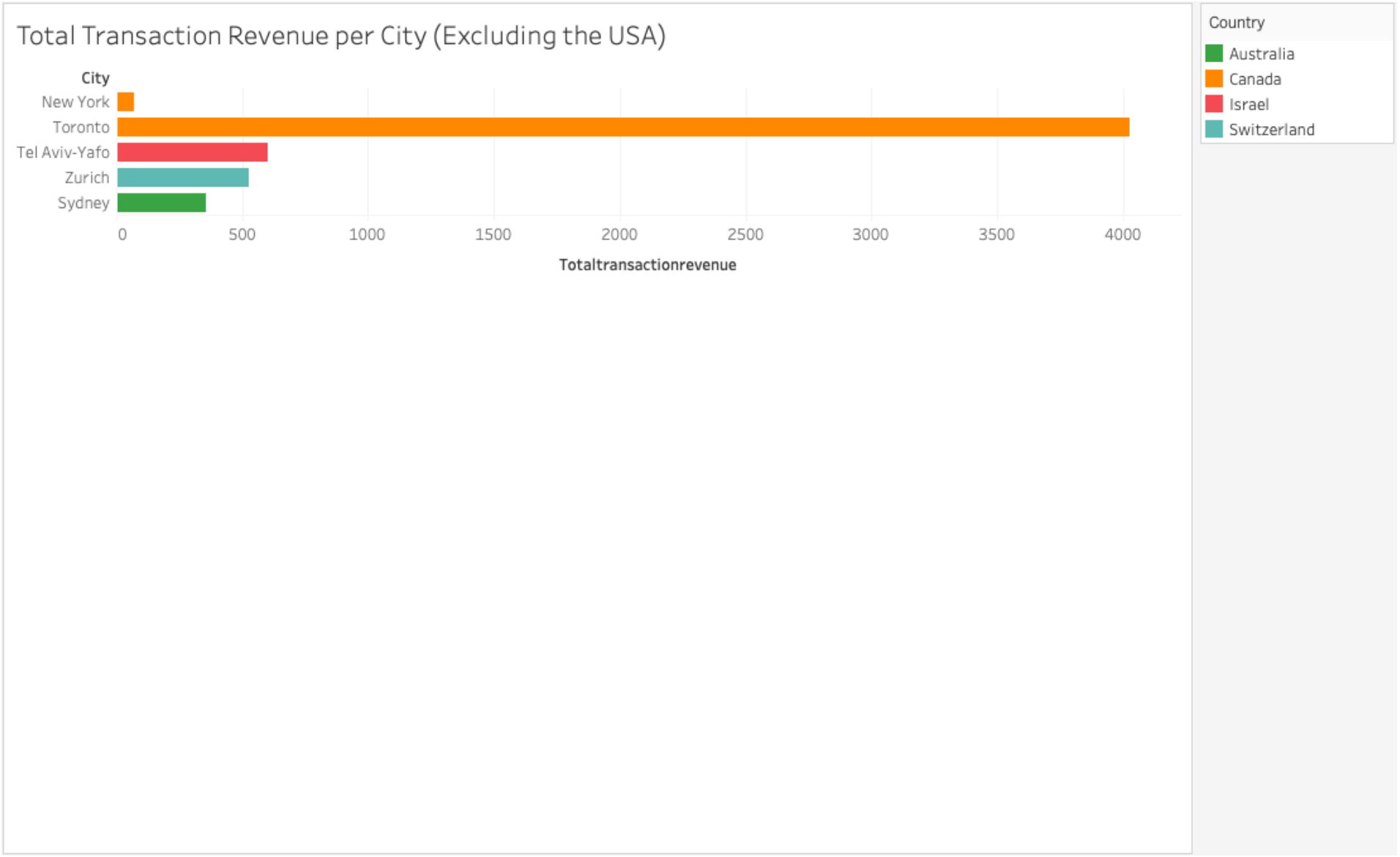
Total Transaction Revenue per City in the USA



Total Transaction Revenue per City in the USA



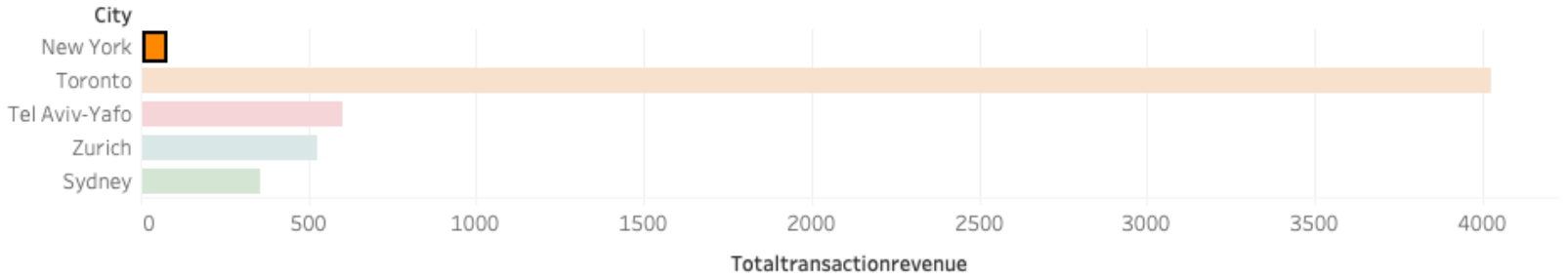

Total Transaction Revenue per Country (Excluding USA)
SUM(Totaltransactionr...
358
4,094
Canada
4,094
Switzerland
527
Israel
602
Australia
358
Russia
Japan
United States
Mexico
Colombia
Peru
Brazil
Argentina
Spain
Turkey
Algeria
Egypt
South Africa
Indonesia
© 2023 Mapbox © OpenStreetMap


Total Transaction Revenue per City (Excluding the USA)
City
New York
Toronto
Tel Aviv-Yafo
Zurich
Sydney
0
500
1000
1500
2000
2500
3000
3500
4000
Totaltransactionrevenue
Country
Australia
Canada
Israel
Switzerland


Total Transaction Revenue per City (Excluding the USA)
City
New York
Toronto
Tel Aviv-Yafo
Zurich
Sydney
0
500
1000
1500
2000
2500
3000
3500
4000
Totaltransactionrevenue
Country
Australia
Canada
Israel
Switzerland

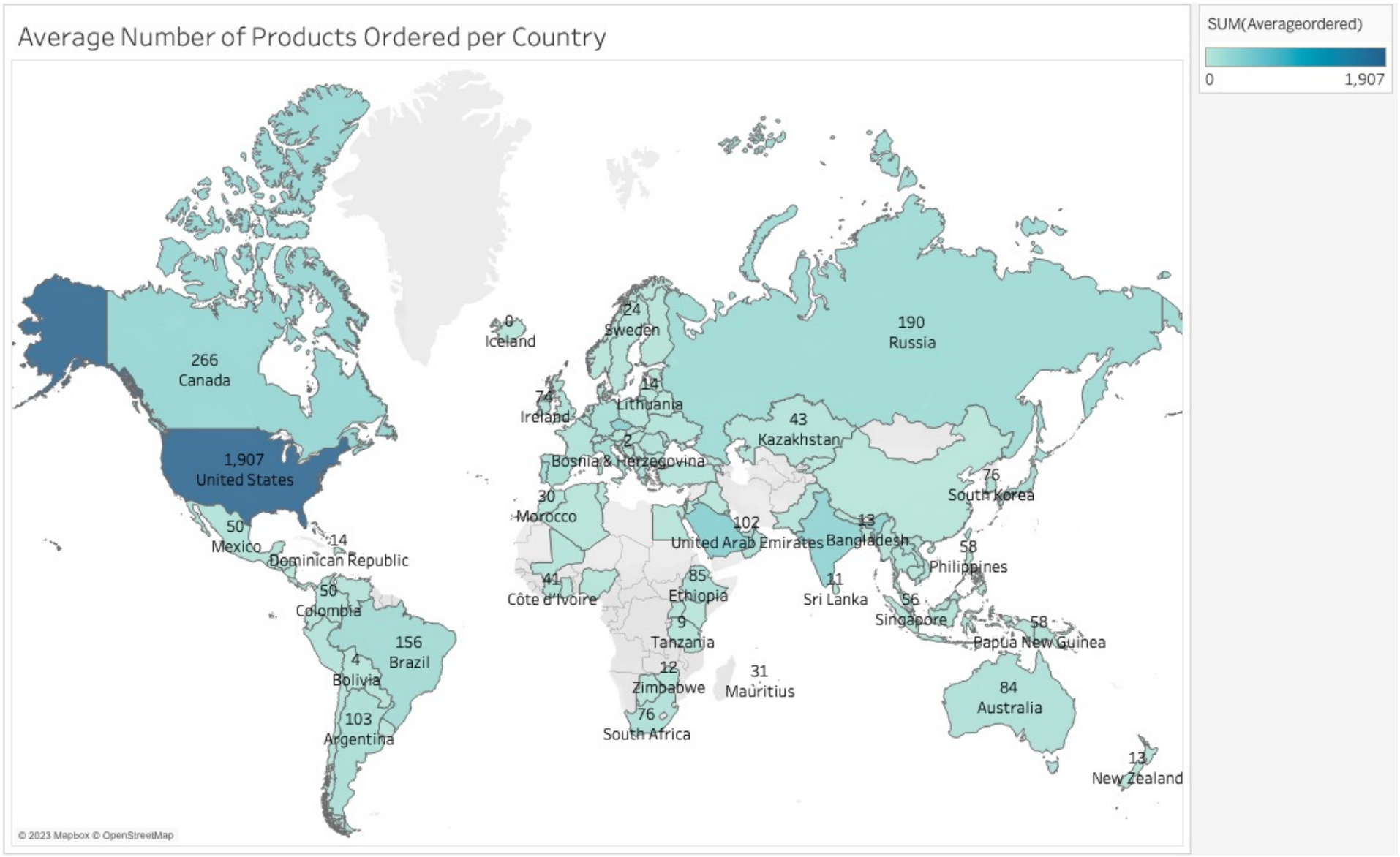

Total Transaction Revenue per City (Excluding the USA)
City
New York
Toronto
Tel Aviv-Yafo
Zurich
Sydney
0
500
1000
1500
2000
2500
3000
3500
4000
Totaltransactionrevenue
Country
Australia
Canada
Israel
Switzerland

Total Transaction Revenue per City (Excluding the USA)

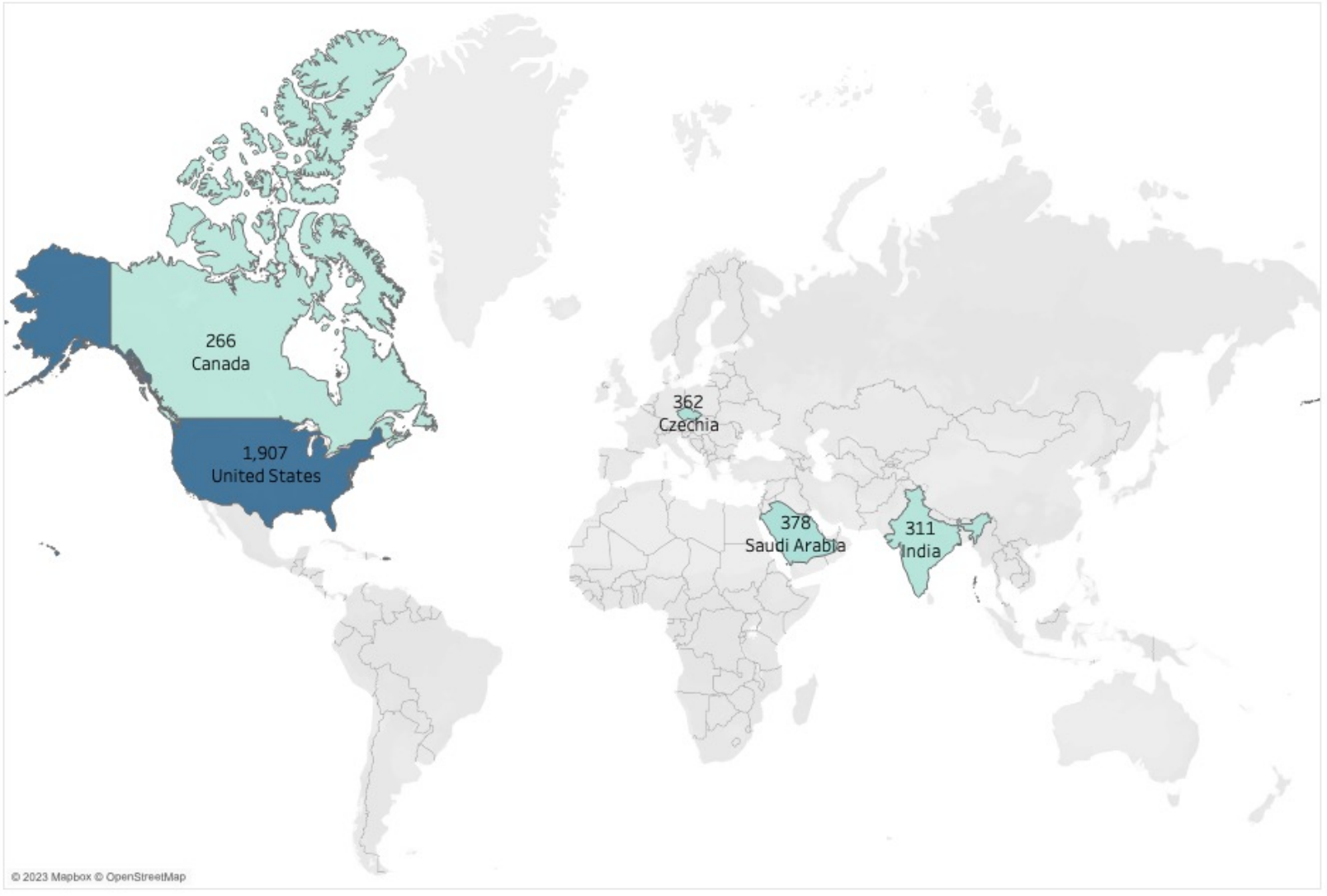


Country
- Australia
- Canada
- Israel
- Switzerland

2. What is the average number of products ordered from visitors in each city and country?

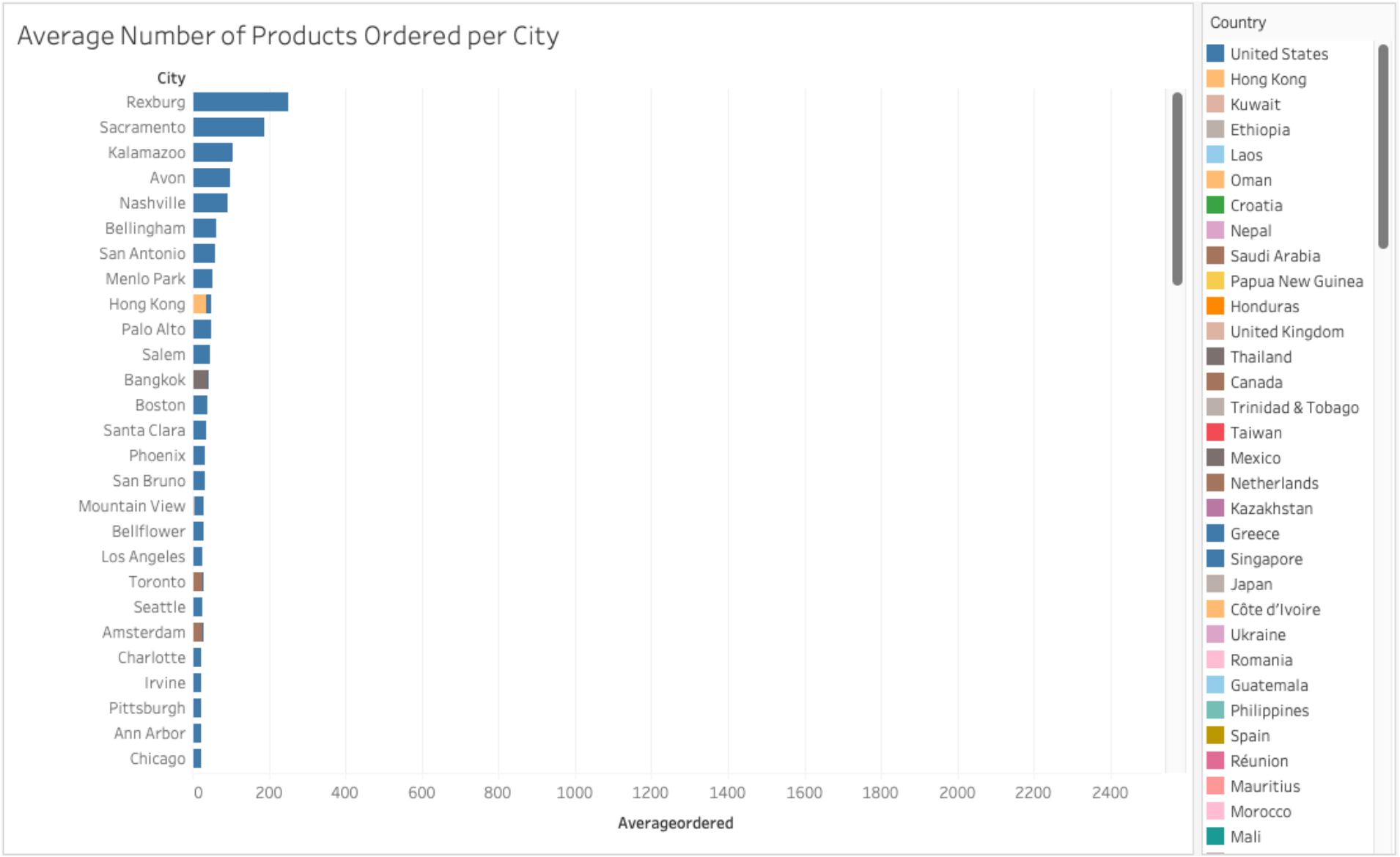

Average Number of Products Ordered per Country
SUM(Averageordered)
0
1,907
266
Canada
1,907
United States
50
Mexico
14
Dominican Republic
50
Colombia
156
Brazil
4
Bolivia
103
Argentina
0
Iceland
24
Sweden
74
Ireland
14
Lithuania
2
Bosnia & Herzegovina
30
Morocco
41
Côte d'Ivoire
85
Ethiopia
9
Tanzania
12
Zimbabwe
76
South Africa
31
Mauritius
190
Russia
43
Kazakhstan
102
United Arab Emirates
13
Bangladesh
11
Sri Lanka
76
South Korea
58
Philippines
56
Singapore
58
Papua New Guinea
84
Australia
13
New Zealand
© 2023 Mapbox © OpenStreetMap

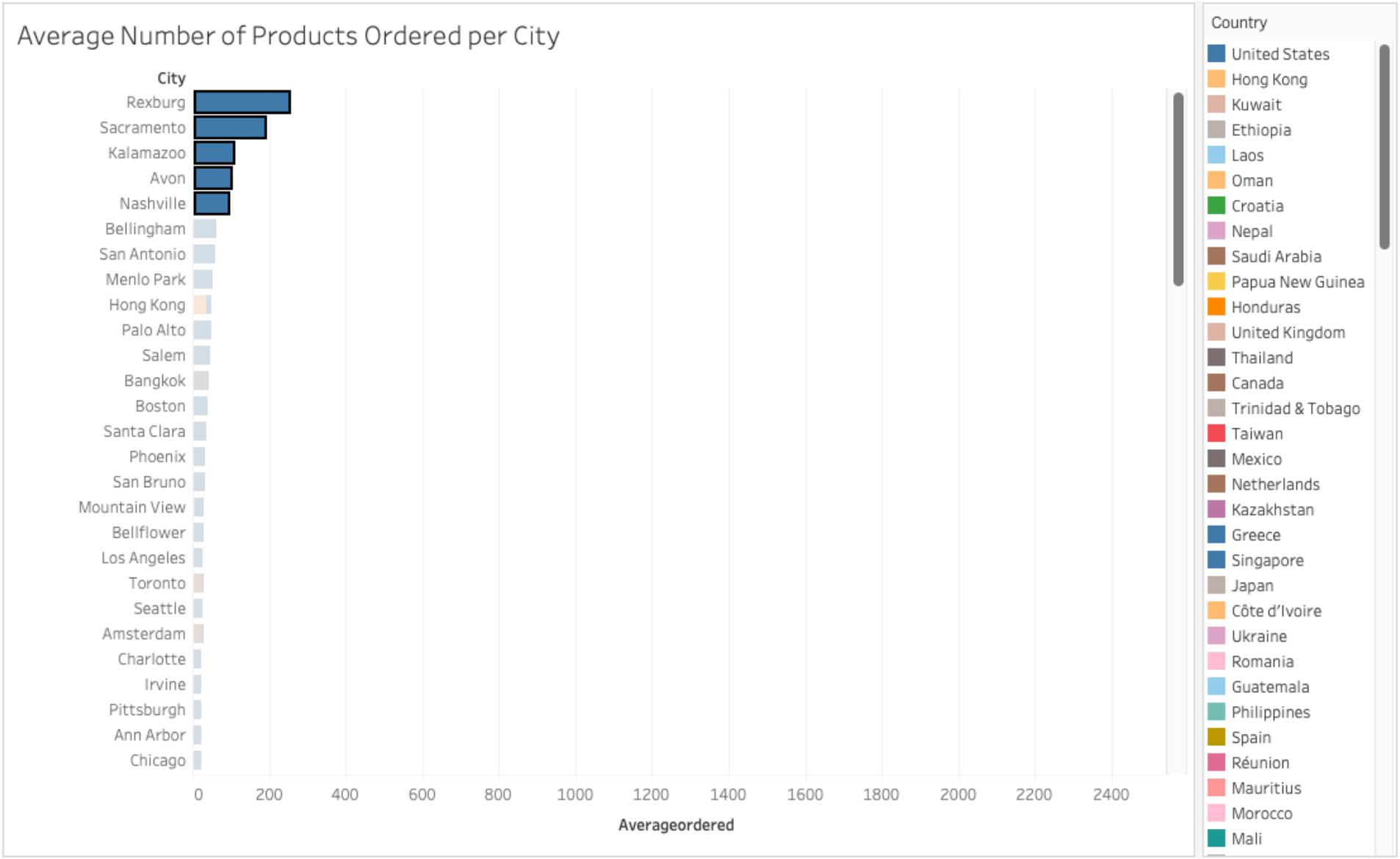
Average Number of Products Ordered per Country

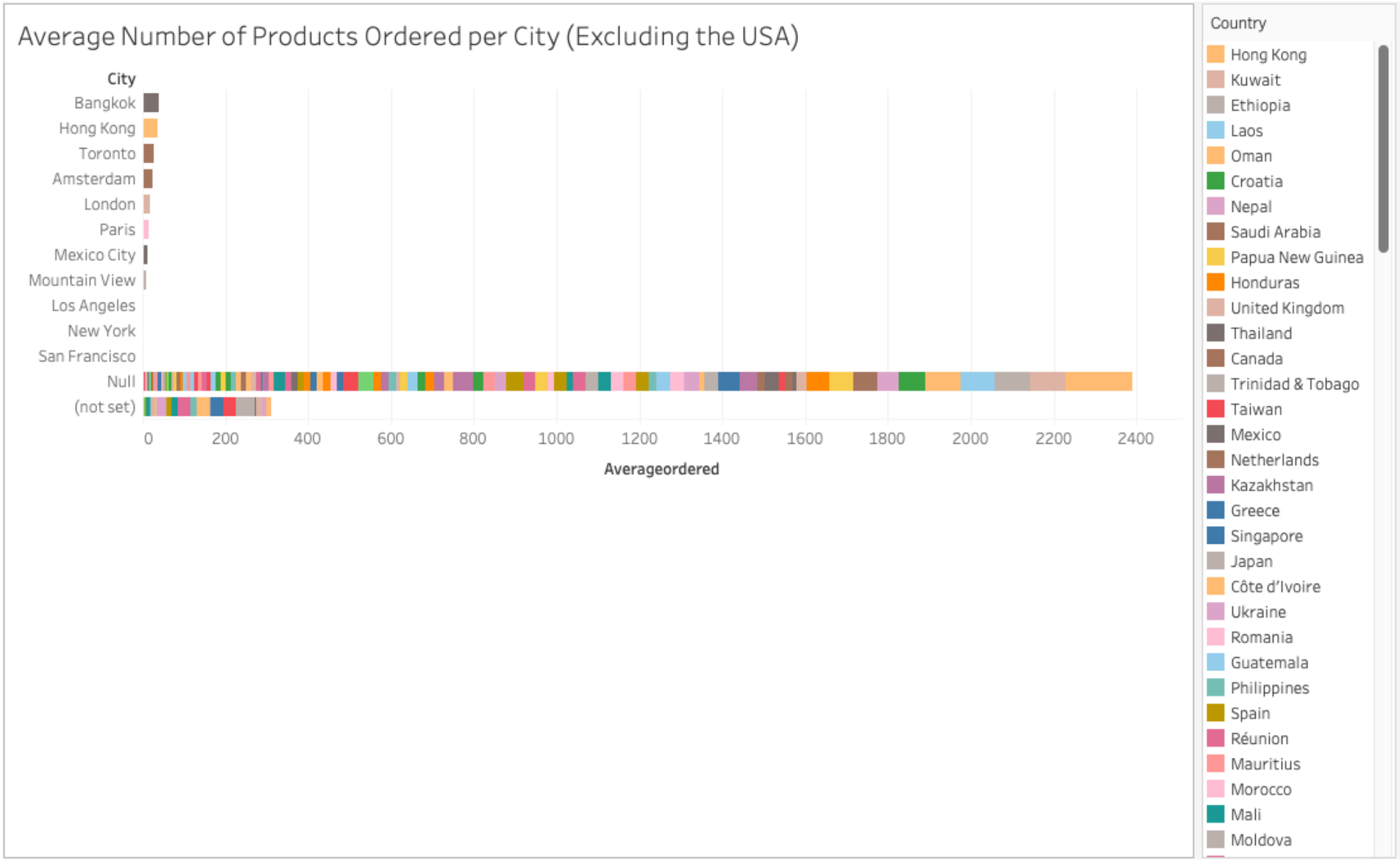


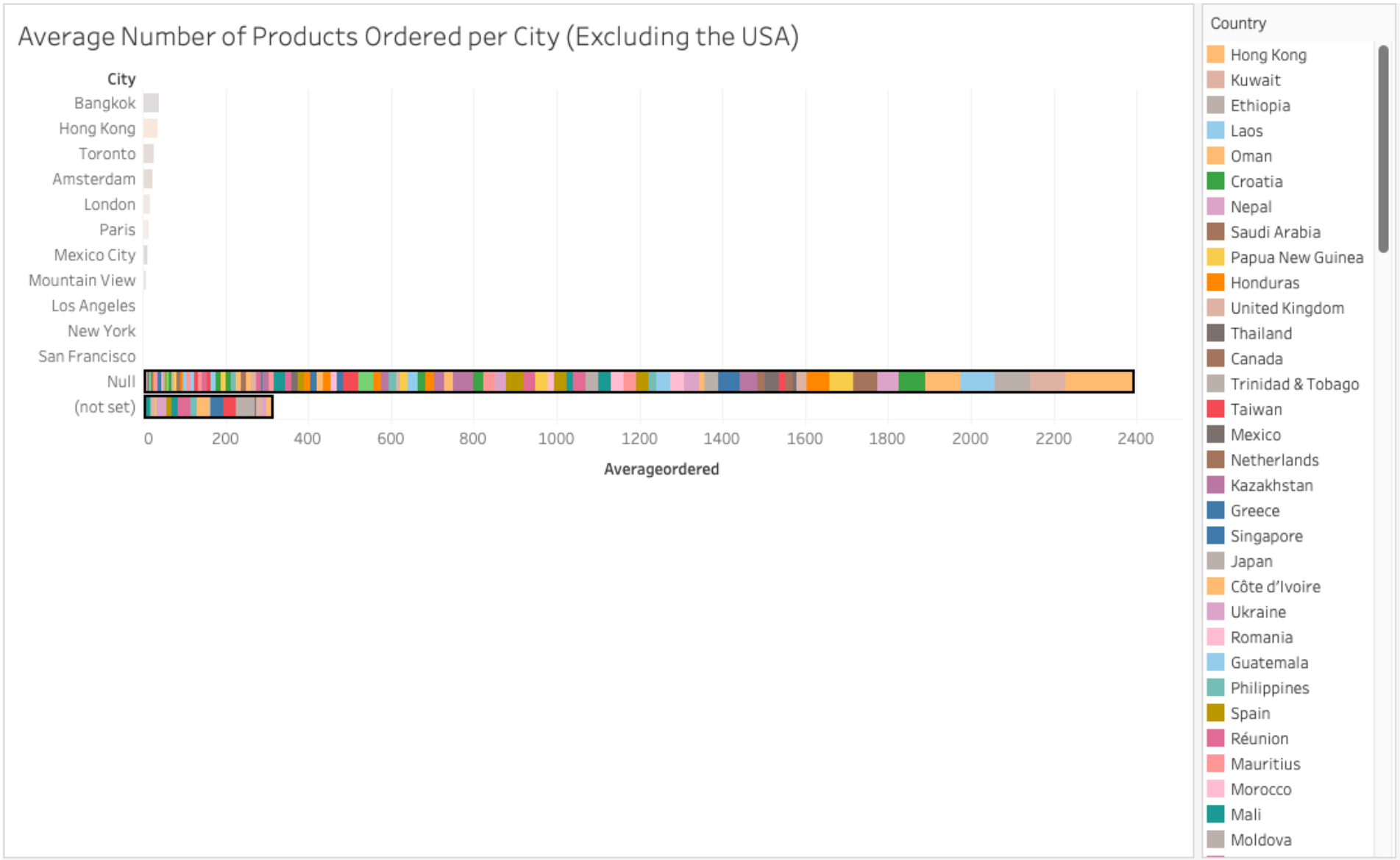
SUM(Averageordered)

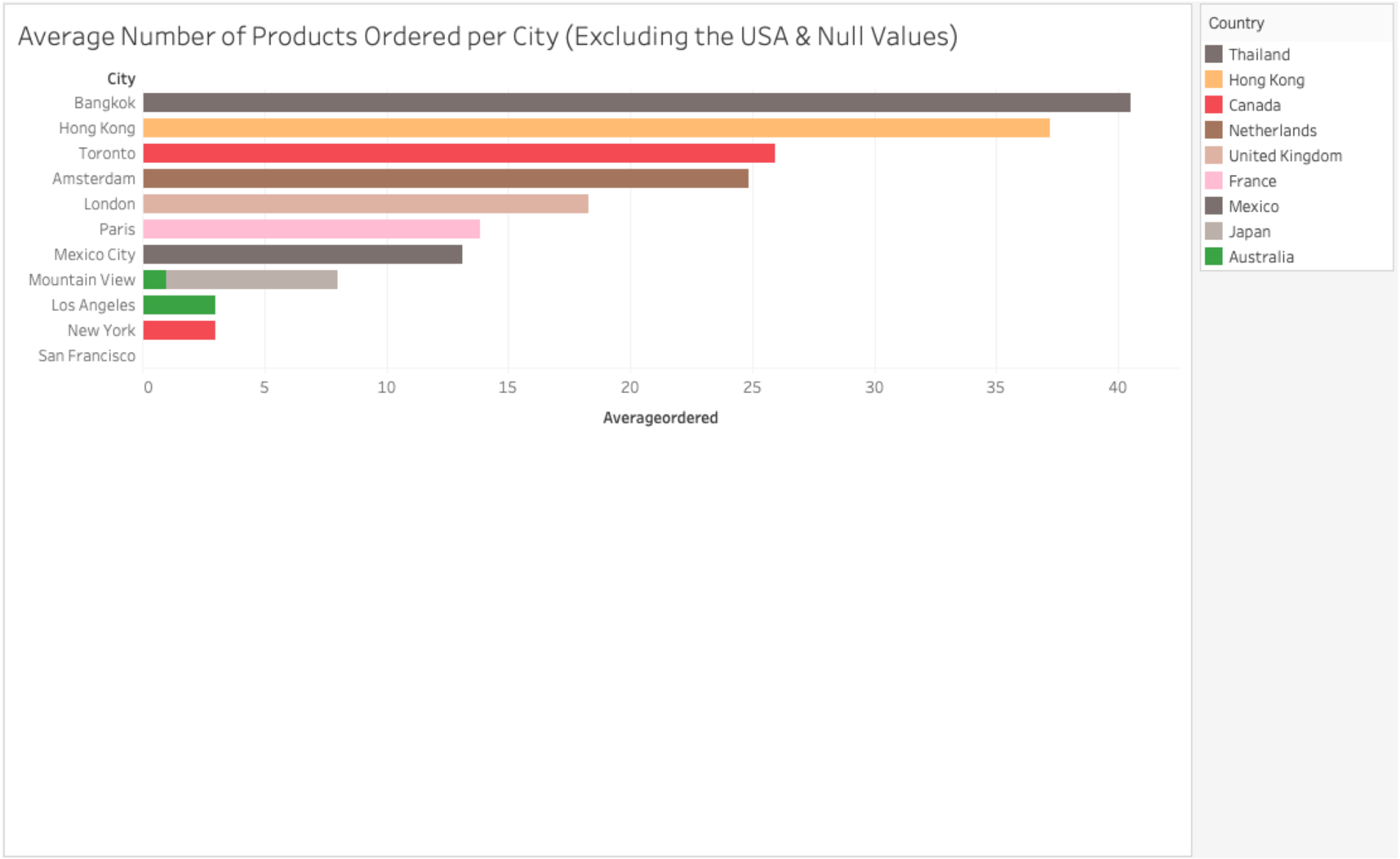
266 1,907

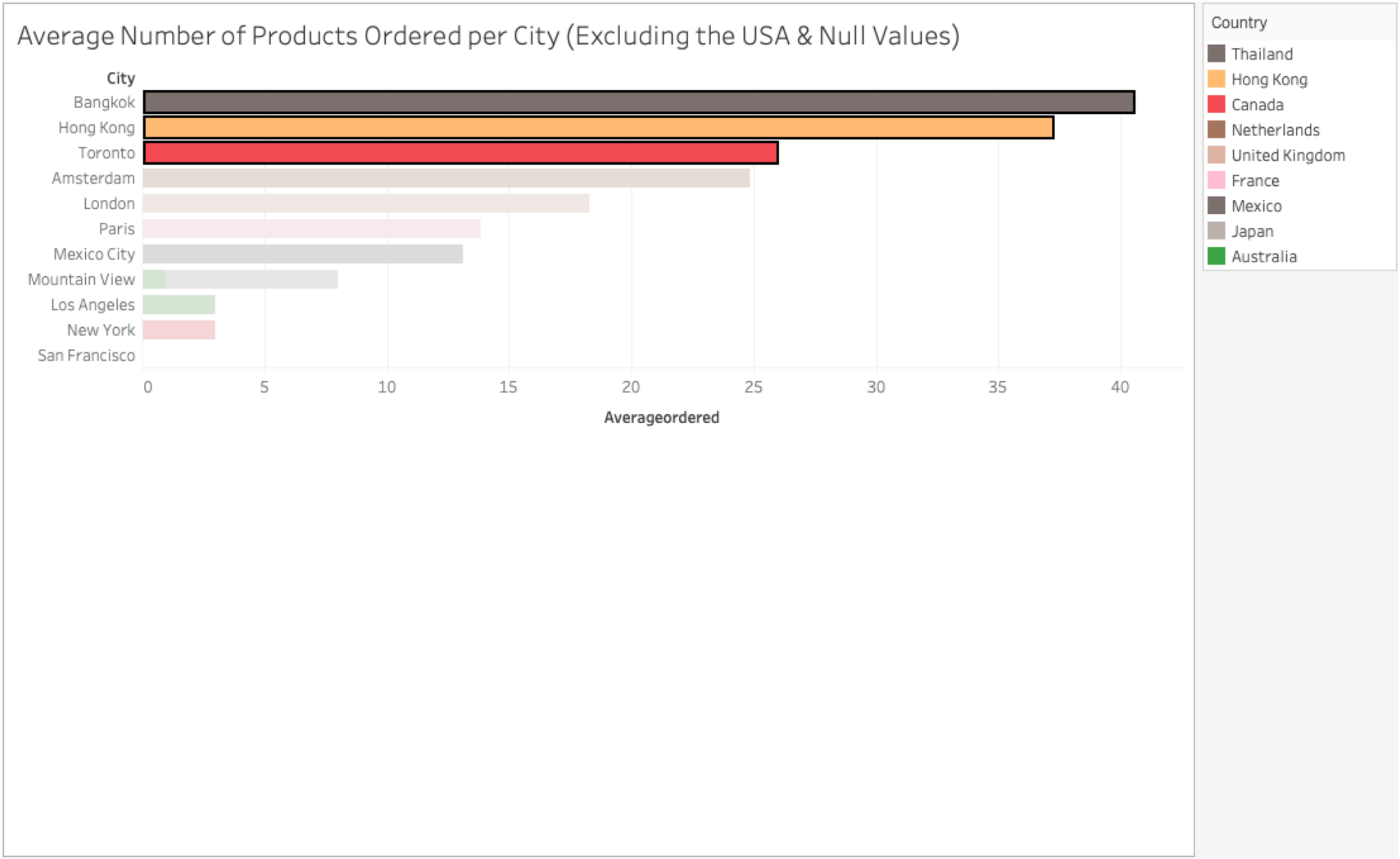

Average Number of Products Ordered per City
City
Rexburg
Sacramento
Kalamazoo
Avon
Nashville
Bellingham
San Antonio
Menlo Park
Hong Kong
Palo Alto
Salem
Bangkok
Boston
Santa Clara
Phoenix
San Bruno
Mountain View
Bellflower
Los Angeles
Toronto
Seattle
Amsterdam
Charlotte
Irvine
Pittsburgh
Ann Arbor
Chicago
0
200
400
600
800
1000
1200
1400
1600
1800
2000
2200
2400
Averageordered
Country
United States
Hong Kong
Kuwait
Ethiopia
Laos
Oman
Croatia
Nepal
Saudi Arabia
Papua New Guinea
Honduras
United Kingdom
Thailand
Canada
Trinidad & Tobago
Taiwan
Mexico
Netherlands
Kazakhstan
Greece
Singapore
Japan
Côte d'Ivoire
Ukraine
Romania
Guatemala
Philippines
Spain
Réunion
Mauritius
Morocco
Mali

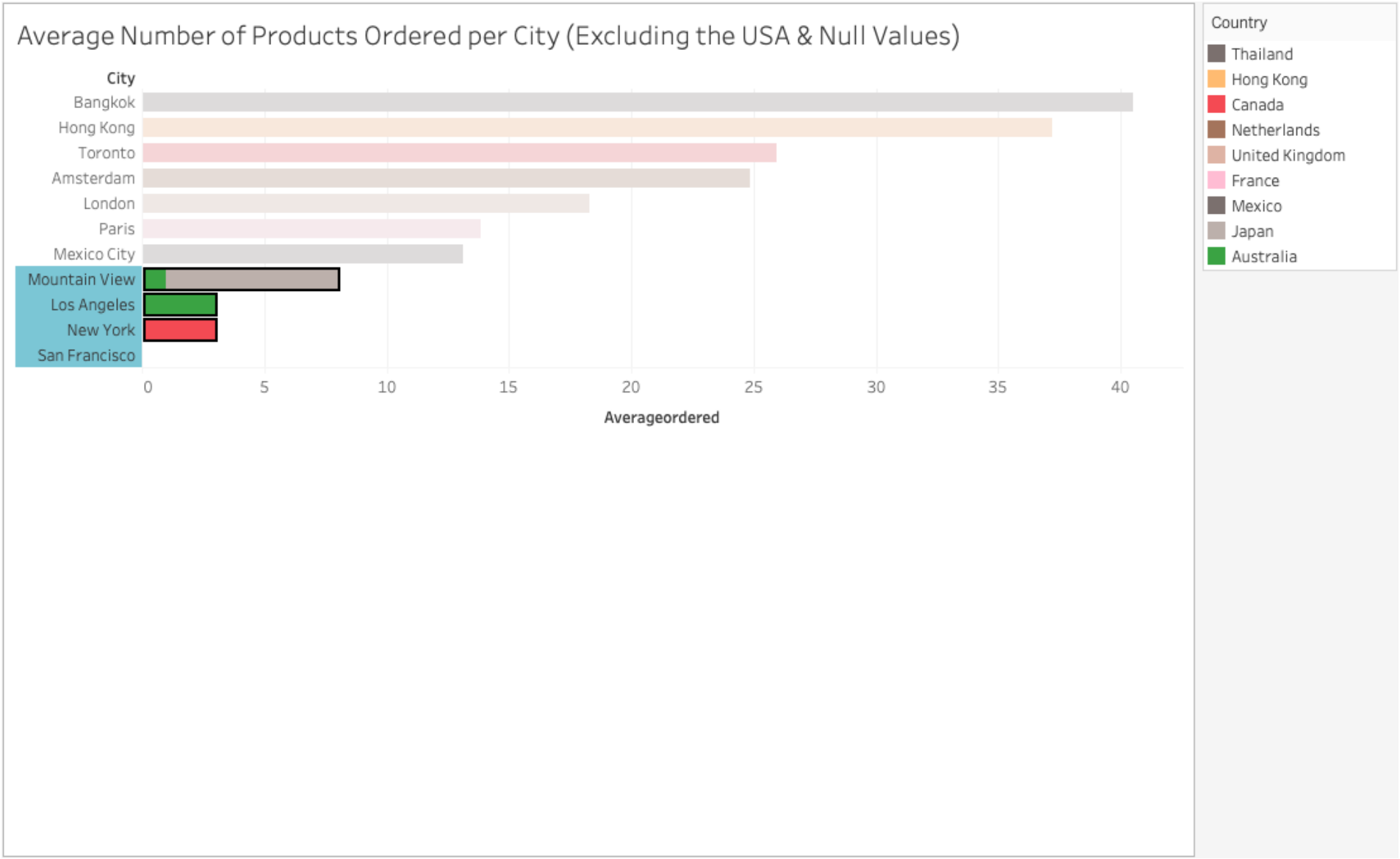

Average Number of Products Ordered per City
City
Rexburg
Sacramento
Kalamazoo
Avon
Nashville
Bellingham
San Antonio
Menlo Park
Hong Kong
Palo Alto
Salem
Bangkok
Boston
Santa Clara
Phoenix
San Bruno
Mountain View
Bellflower
Los Angeles
Toronto
Seattle
Amsterdam
Charlotte
Irvine
Pittsburgh
Ann Arbor
Chicago
0
200
400
600
800
1000
1200
1400
1600
1800
2000
2200
2400
Averageordered
Country
United States
Hong Kong
Kuwait
Ethiopia
Laos
Oman
Croatia
Nepal
Saudi Arabia
Papua New Guinea
Honduras
United Kingdom
Thailand
Canada
Trinidad & Tobago
Taiwan
Mexico
Netherlands
Kazakhstan
Greece
Singapore
Japan
Côte d'Ivoire
Ukraine
Romania
Guatemala
Philippines
Spain
Réunion
Mauritius
Morocco
Mali


Average Number of Products Ordered per City (Excluding the USA)
City
Bangkok
Hong Kong
Toronto
Amsterdam
London
Paris
Mexico City
Mountain View
Los Angeles
New York
San Francisco
Null
(not set)
0
200
400
600
800
1000
1200
1400
1600
1800
2000
2200
2400
Averageordered
Country
Hong Kong
Kuwait
Ethiopia
Laos
Oman
Croatia
Nepal
Saudi Arabia
Papua New Guinea
Honduras
United Kingdom
Thailand
Canada
Trinidad & Tobago
Taiwan
Mexico
Netherlands
Kazakhstan
Greece
Singapore
Japan
Côte d'Ivoire
Ukraine
Romania
Guatemala
Philippines
Spain
Réunion
Mauritius
Morocco
Mali
Moldova


Average Number of Products Ordered per City (Excluding the USA)
City
Bangkok
Hong Kong
Toronto
Amsterdam
London
Paris
Mexico City
Mountain View
Los Angeles
New York
San Francisco
Null
(not set)
0
200
400
600
800
1000
1200
1400
1600
1800
2000
2200
2400
Averageordered
Country
Hong Kong
Kuwait
Ethiopia
Laos
Oman
Croatia
Nepal
Saudi Arabia
Papua New Guinea
Honduras
United Kingdom
Thailand
Canada
Trinidad & Tobago
Taiwan
Mexico
Netherlands
Kazakhstan
Greece
Singapore
Japan
Côte d'Ivoire
Ukraine
Romania
Guatemala
Philippines
Spain
Réunion
Mauritius
Morocco
Mali
Moldova


Average Number of Products Ordered per City (Excluding the USA & Null Values)
City
Bangkok
Hong Kong
Toronto
Amsterdam
London
Paris
Mexico City
Mountain View
Los Angeles
New York
San Francisco
0
5
10
15
20
25
30
35
40
Averageordered
Country
Thailand
Hong Kong
Canada
Netherlands
United Kingdom
France
Mexico
Japan
Australia

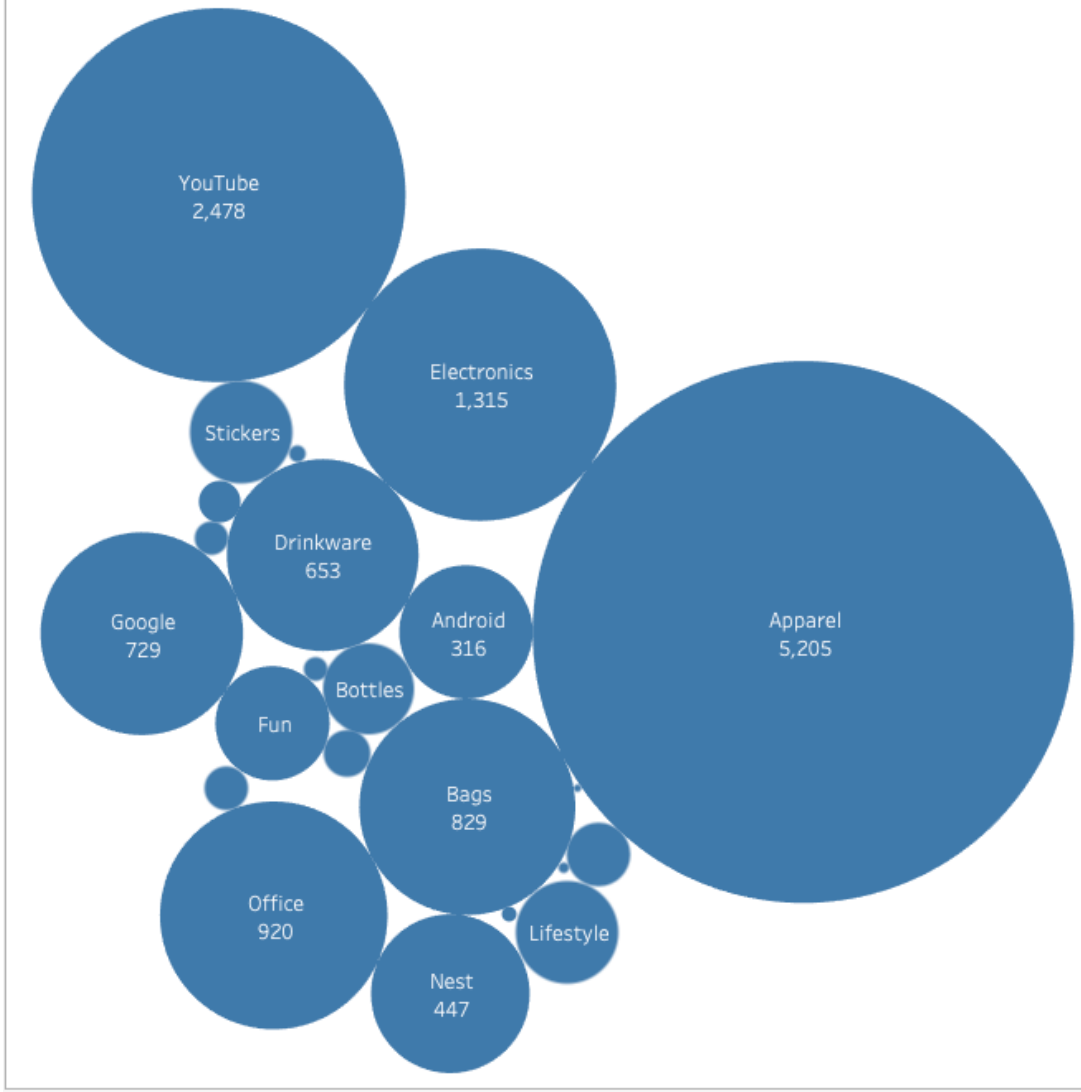

Average Number of Products Ordered per City (Excluding the USA & Null Values)
City
Bangkok
Hong Kong
Toronto
Amsterdam
London
Paris
Mexico City
Mountain View
Los Angeles
New York
San Francisco
0
5
10
15
20
25
30
35
40
Averageordered
Country
Thailand
Hong Kong
Canada
Netherlands
United Kingdom
France
Mexico
Japan
Australia


Average Number of Products Ordered per City (Excluding the USA & Null Values)
City
Bangkok
Hong Kong
Toronto
Amsterdam
London
Paris
Mexico City
Mountain View
Los Angeles
New York
San Francisco
0
5
10
15
20
25
30
35
40
Averageordered
Country
Thailand
Hong Kong
Canada
Netherlands
United Kingdom
France
Mexico
Japan
Australia

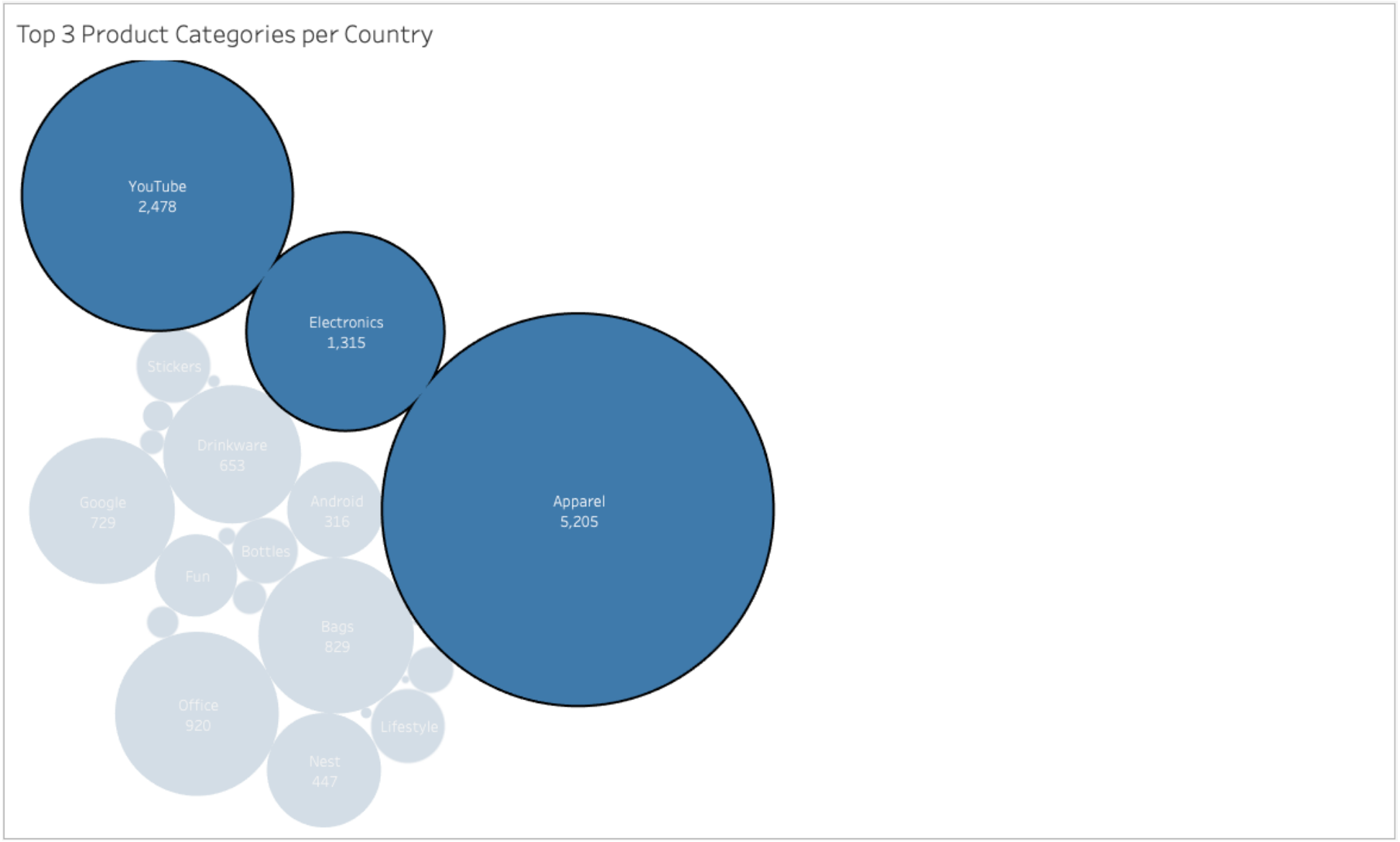
3. Is there any pattern in the types (product categories) of products ordered from visitors in each city and country?


Product Count Distribution per Country
SUM(Productcatcount)
1
4,832
Iceland
1
Russia
65
Canada
404
United States
4,832
Mexico
61
Haiti
1
Brazil
88
Argentina
24
Polynesia
1
Greece
26
Morocco
5
Jordan
1
Kyrgyzstan
1
China
15
Japan
136
Pakistan
38
Macau
1
Sudan
2
Côte d'Ivoire
4
Somalia
1
Tanzania
1
Zimbabwe
1
South Africa
21
Réunion
1
Maldives
1
Indonesia
64
Papua New Guinea
2
Australia
152
New Zeala
40
© 2023 Mapbox © OpenStreetMap

Top 3 Product Categories per Country



Top 3 Product Categories per Country

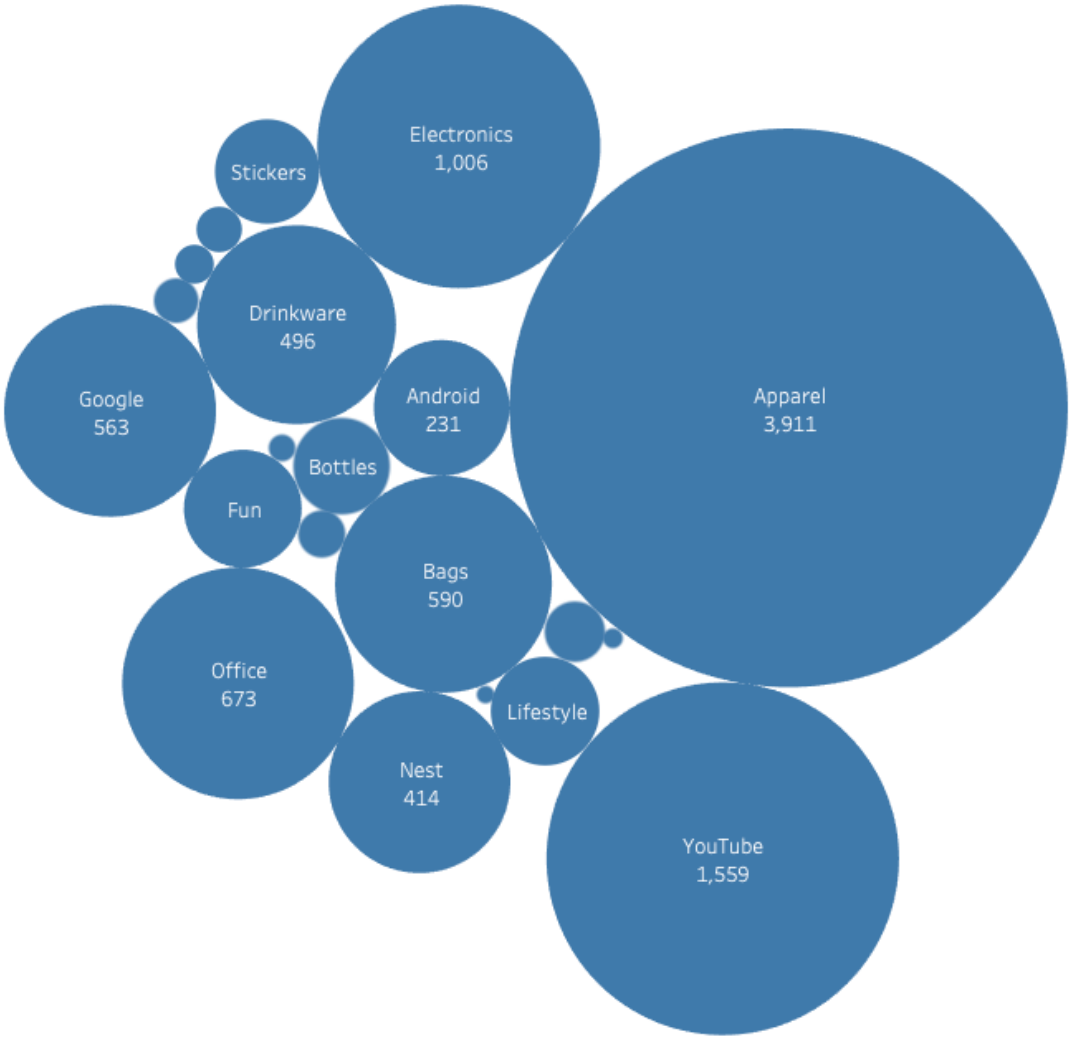


Product Count Distribution per Country

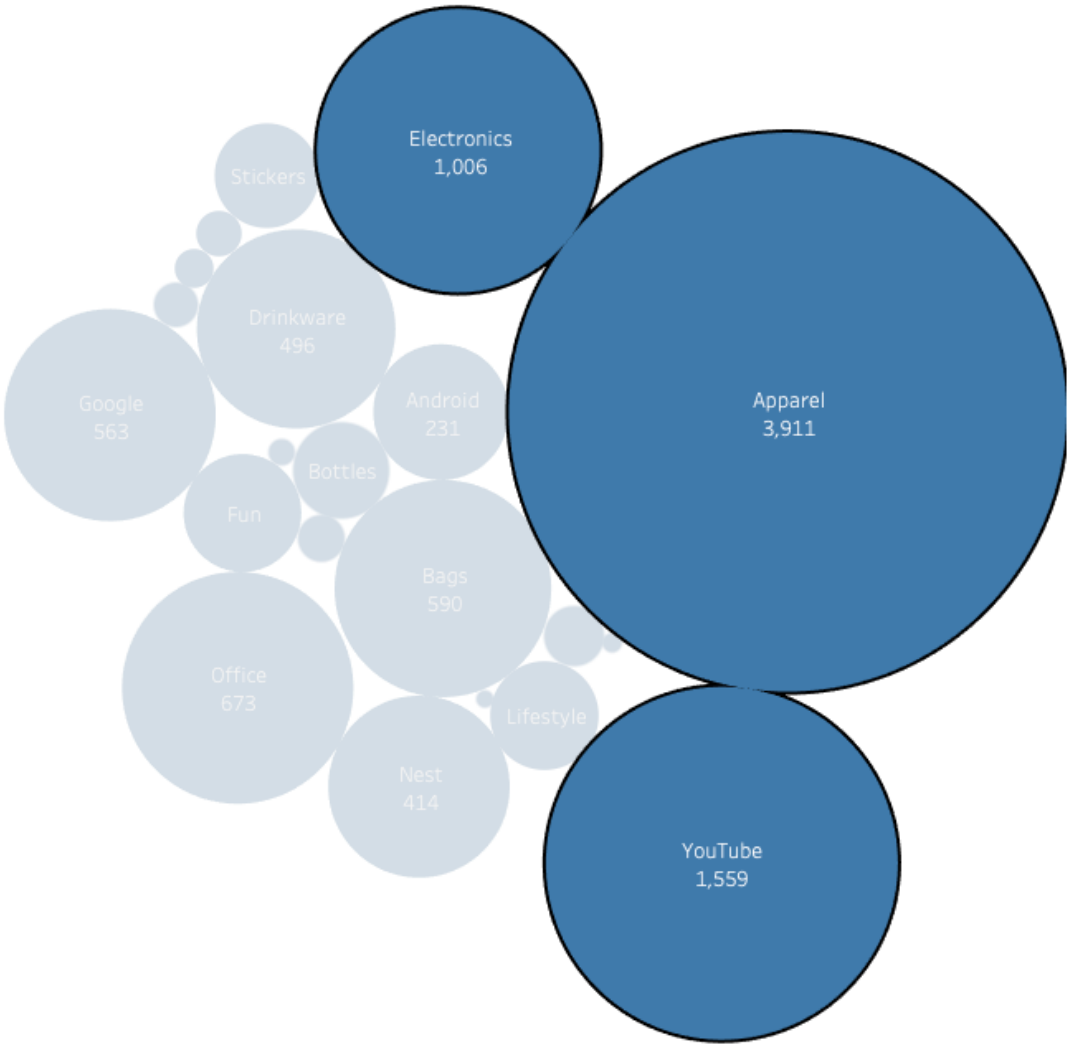
SUM(Productcatcount)

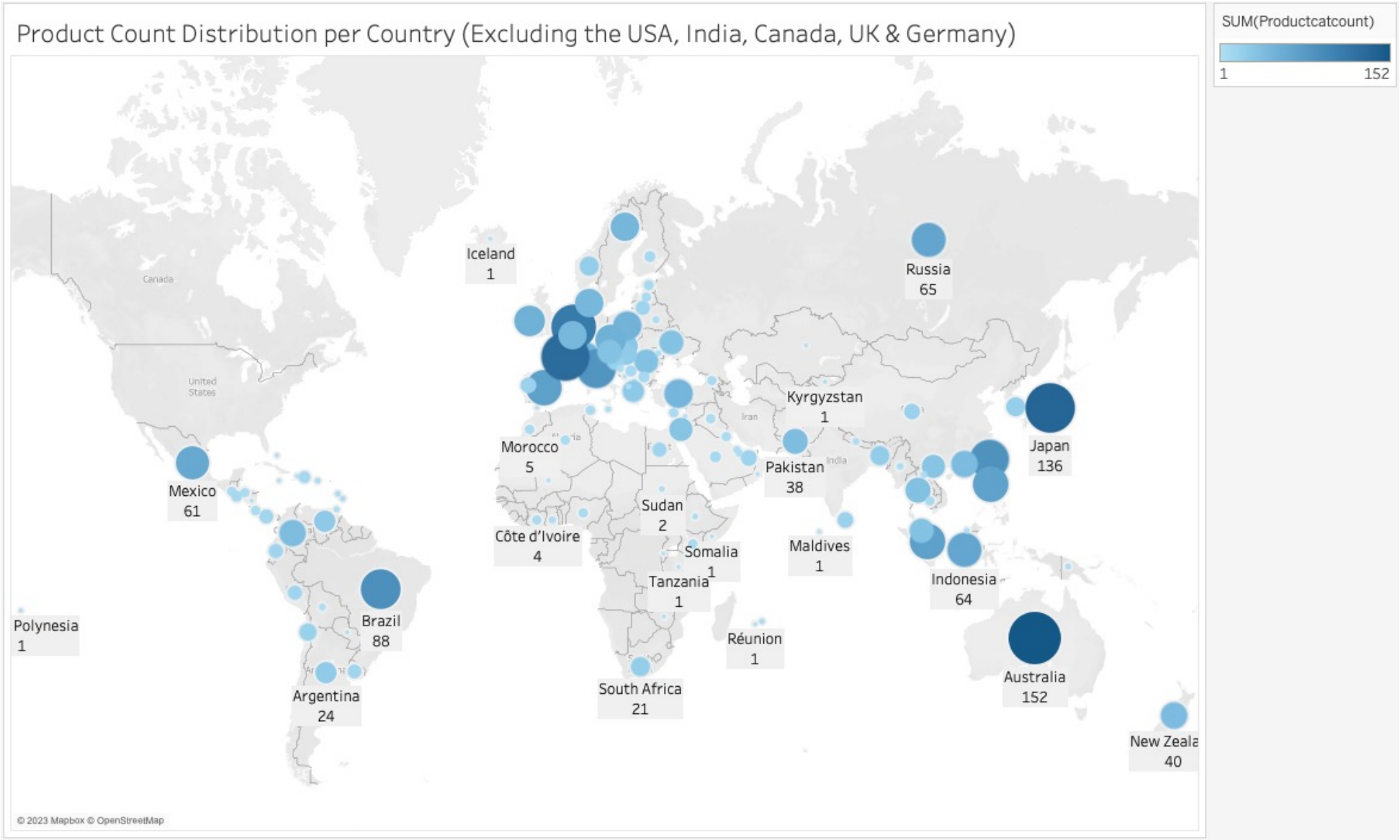
253 4,832

Top 3 Product Categories per Country on the Top 5 Countries (USA, India, Canada, UK and Germany)

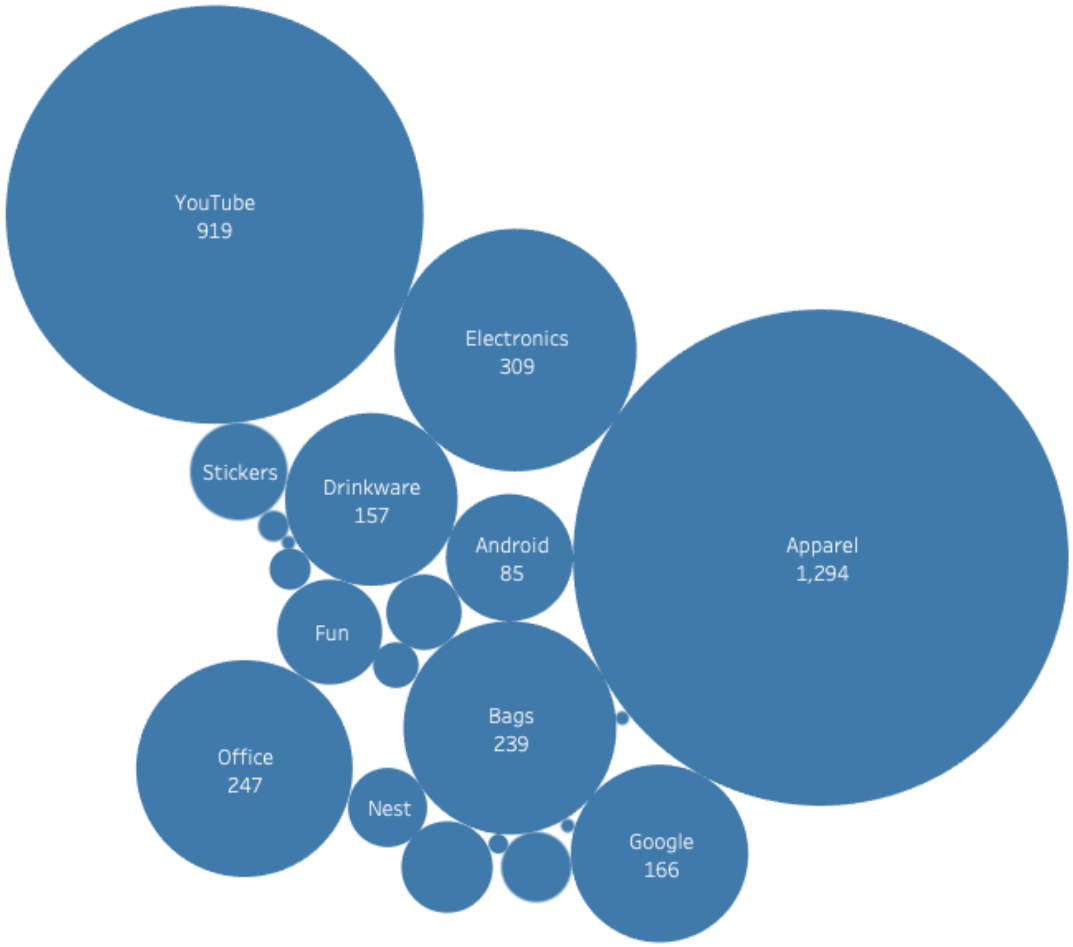


Top 3 Product Categories per Country on the Top 5 Countries (USA, India, Canada, UK and Germany)

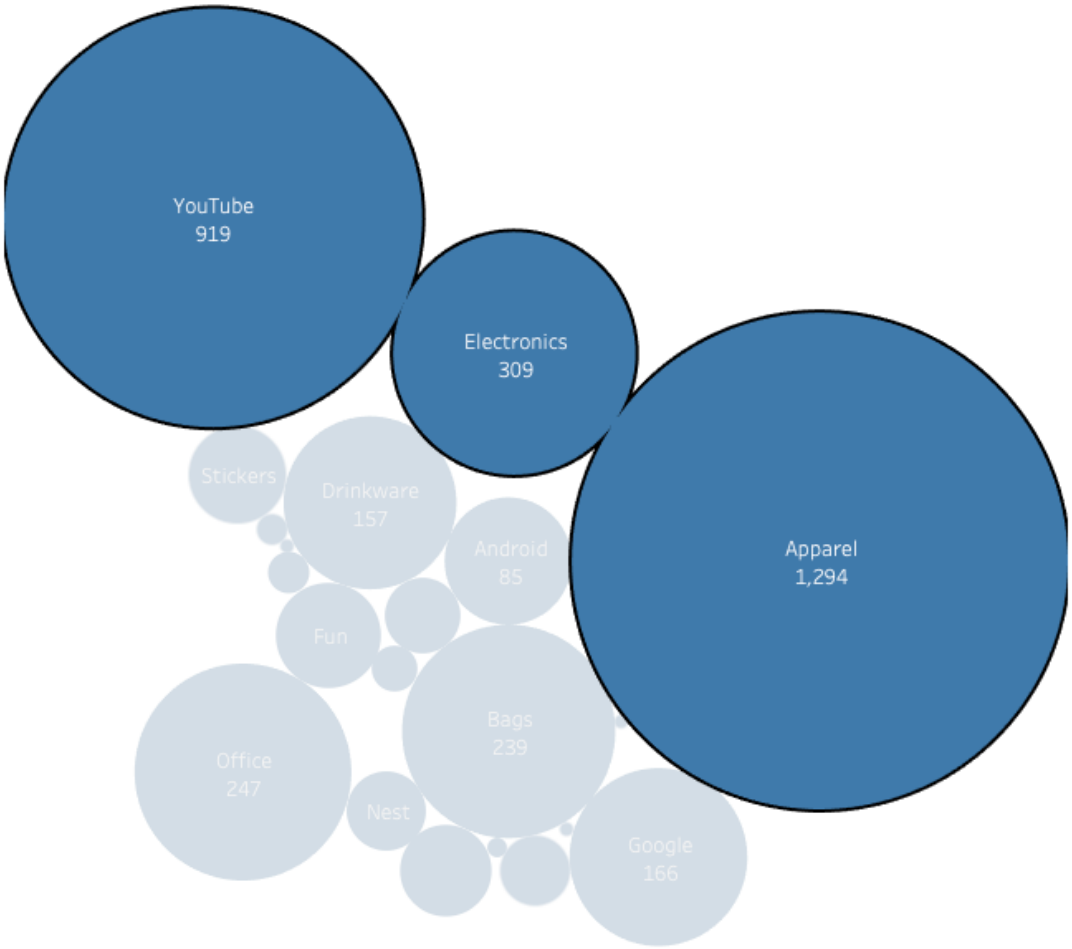



Product Count Distribution per Country (Excluding the USA, India, Canada, UK & Germany)
SUM(Productcatcount)
1
152
Iceland
1
Russia
65
Kyrgyzstan
1
Japan
136
Morocco
5
Pakistan
38
Mexico
61
Sudan
2
Côte d'Ivoire
4
Somalia
1
Maldives
1
Tanzania
1
Indonesia
64
Polynesia
1
Brazil
88
Réunion
1
Australia
152
Argentina
24
South Africa
21
New Zeala
40
Canada
United States
Iran
India
© 2023 Mapbox © OpenStreetMap

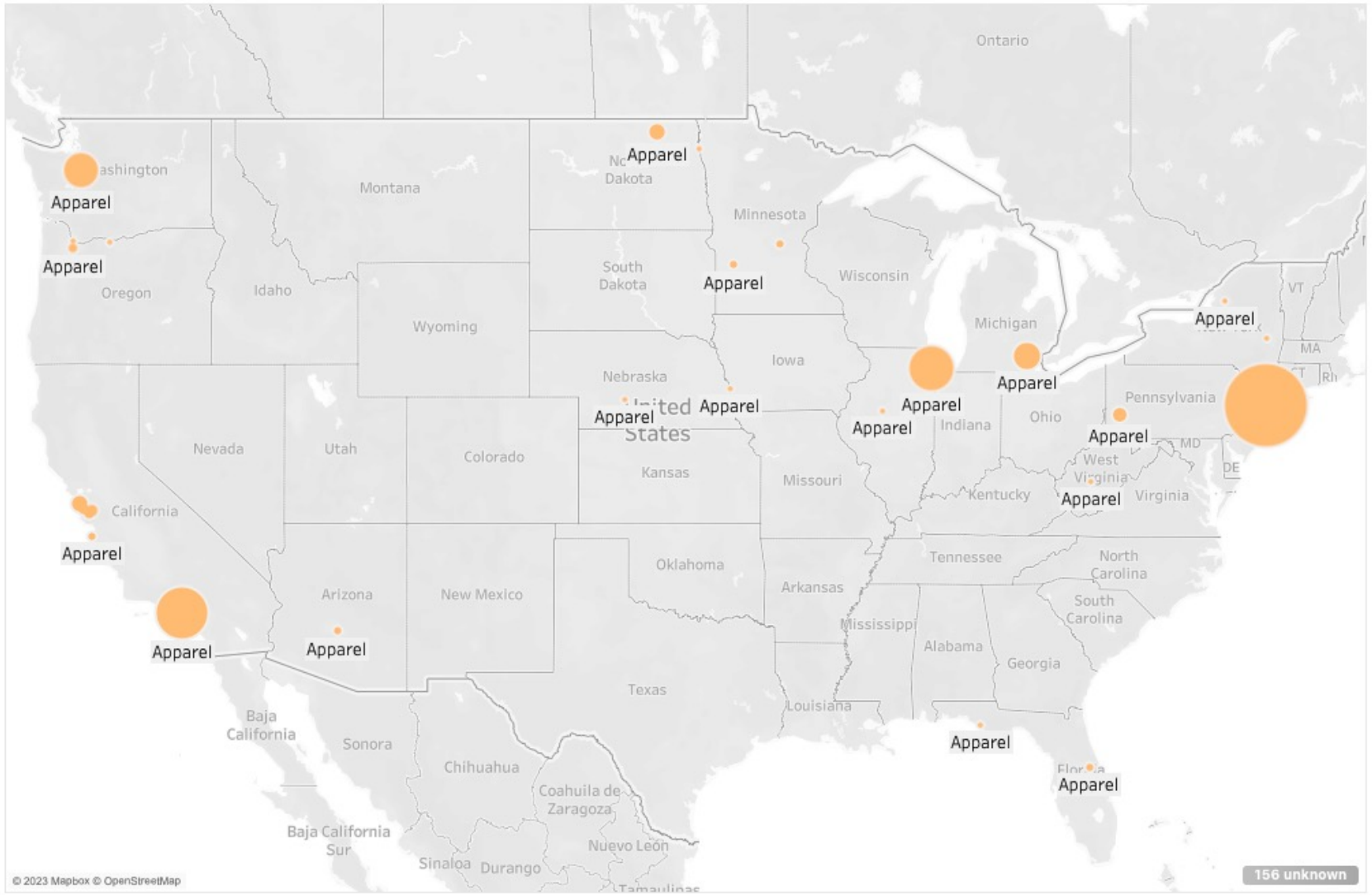
Top 3 Product Categories per Country (Excluding USA, India, Canada, UK and Germany)



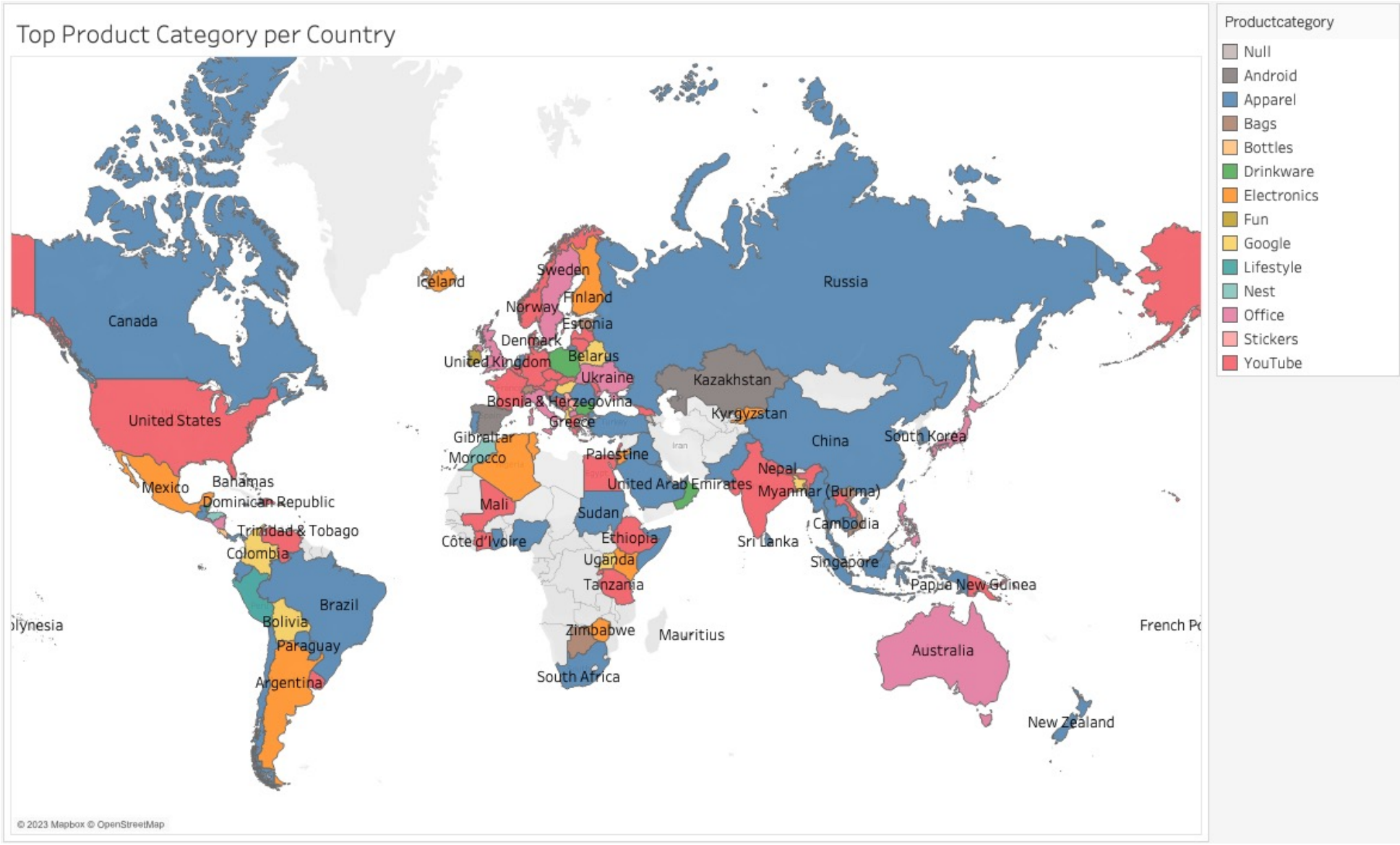
Top 3 Product Categories per Country (Excluding USA, India, Canada, UK and Germany)

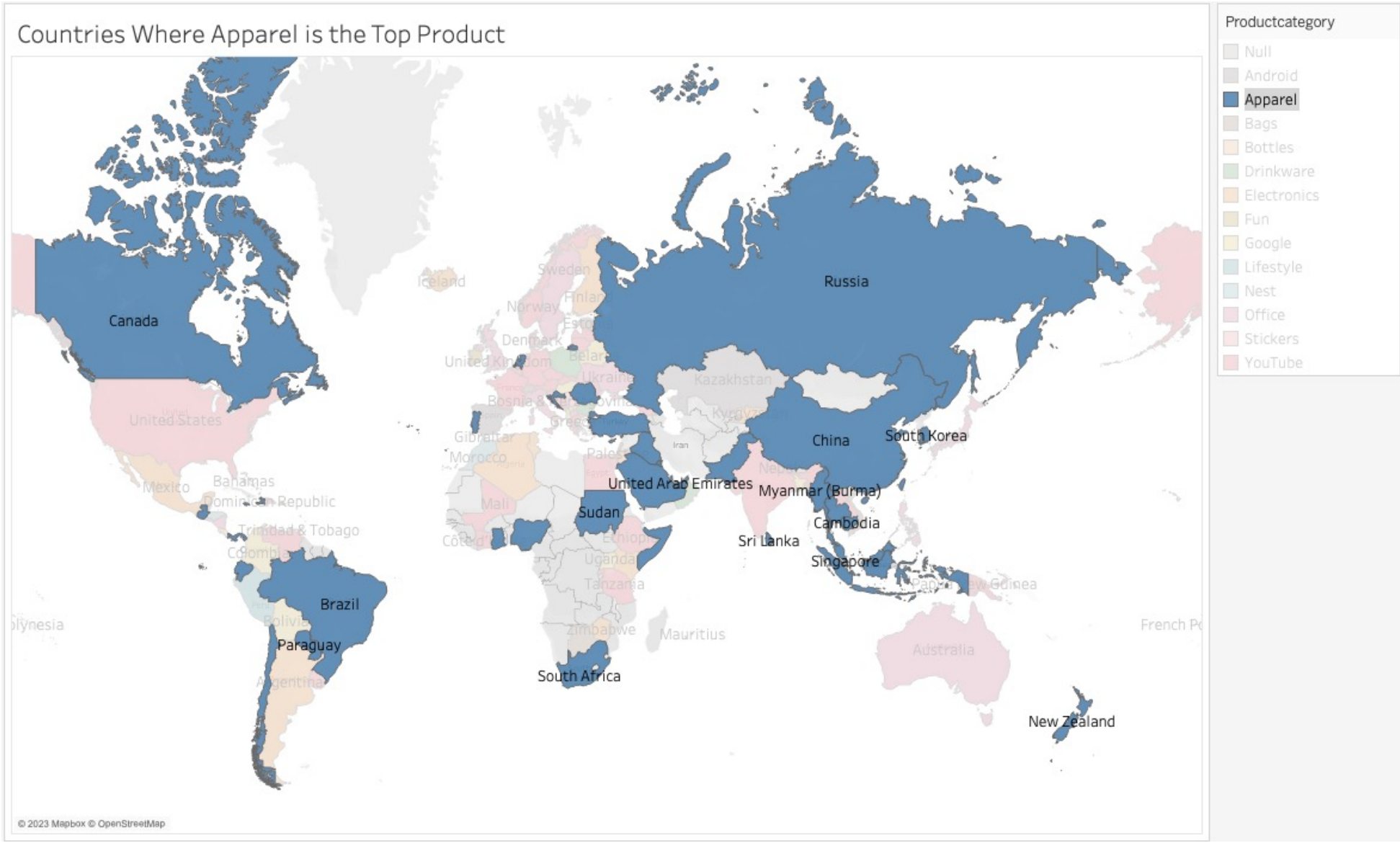


Top Product Category per City (Only USA is Available)



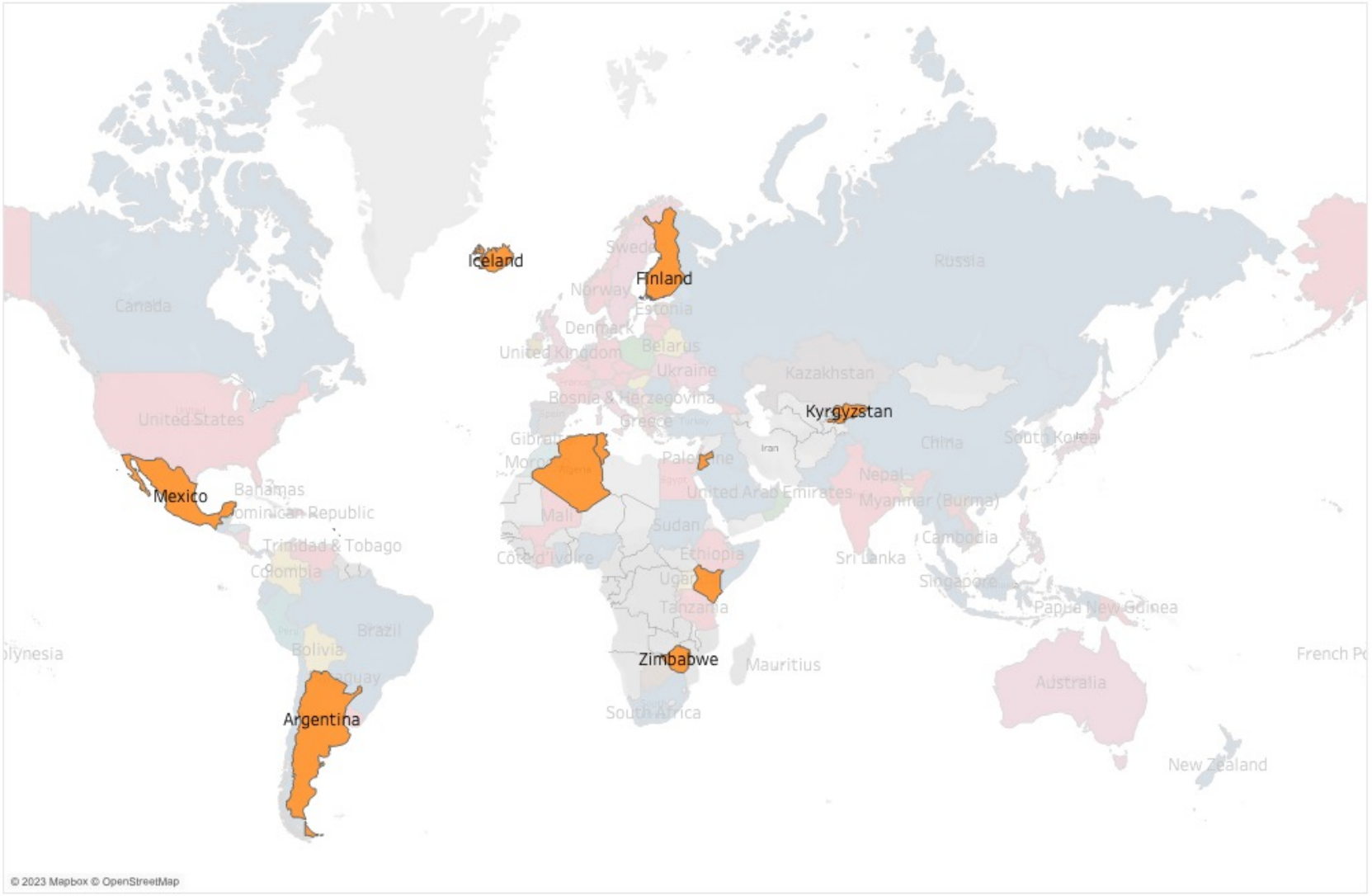
4. What is the top-selling product from each city/country? Can we find any pattern worthy of noting in the products sold?


Top Product Category per Country
Productcategory
Null
Android
Apparel
Bags
Bottles
Drinkware
Electronics
Fun
Google
Lifestyle
Nest
Office
Stickers
YouTube
Canada
United States
Mexico
Bahamas
Dominican Republic
Trinidad & Tobago
Colombia
Brazil
Bolivia
Paraguay
Argentina
Iceland
Sweden
Norway
Finland
Estonia
Denmark
United Kingdom
Belarus
Ukraine
Bosnia & Herzegovina
Greece
Gibraltar
Morocco
Mali
Côte d'Ivoire
Palestine
United Arab Emirates
Sudan
Ethiopia
Uganda
Tanzania
Zimbabwe
South Africa
Mauritius
Russia
Kazakhstan
Kyrgyzstan
China
South Korea
Nepal
Myanmar (Burma)
Cambodia
Sri Lanka
Singapore
Papua New Guinea
Australia
New Zealand
© 2023 Mapbox © OpenStreetMap


Countries Where Apparel is the Top Product
Productcategory
Null
Android
Apparel
Bags
Bottles
Drinkware
Electronics
Fun
Google
Lifestyle
Nest
Office
Stickers
YouTube
Canada
Russia
China
South Korea
United Arab Emirates
Myanmar (Burma)
Sudan
Cambodia
Sri Lanka
Singapore
Brazil
Paraguay
South Africa
New Zealand
Iceland
Sweden
Norway
Denmark
United States
Mexico
Bahamas
Trinidad & Tobago
Colombia
Bolivia
Kazakhstan
Iran
Morocco
Mali
Uganda
Tanzania
Zimbabwe
Mauritius
Australia
© 2023 Mapbox © OpenStreetMap


Countries Where Youtube is the Top Product
Productcategory
Null
Android
Apparel
Bags
Bottles
Drinkware
Electronics
Fun
Google
Lifestyle
Nest
Office
Stickers
YouTube
Norway
Denmark
Bosnia & Herzegovina
Greece
Gibraltar
United States
Dominican Republic
Trinidad & Tobago
Mali
Côte d'Ivoire
Ethiopia
Tanzania
Nepal
Papua New Guinea
Mauritius
French Po
Iceland
Canada
Russia
Finland
Kazakhstan
China
South Korea
Mexico
Bahamas
Brazil
Bolivia
Paraguay
Australia
New Zealand
South Africa
Zimbabwe
Sri Lanka
Cambodia
Singapore
Morocco
Iran
Sudan
Uganda
© 2023 Mapbox © OpenStreetMap

Countries Where Electronics is the Top Product

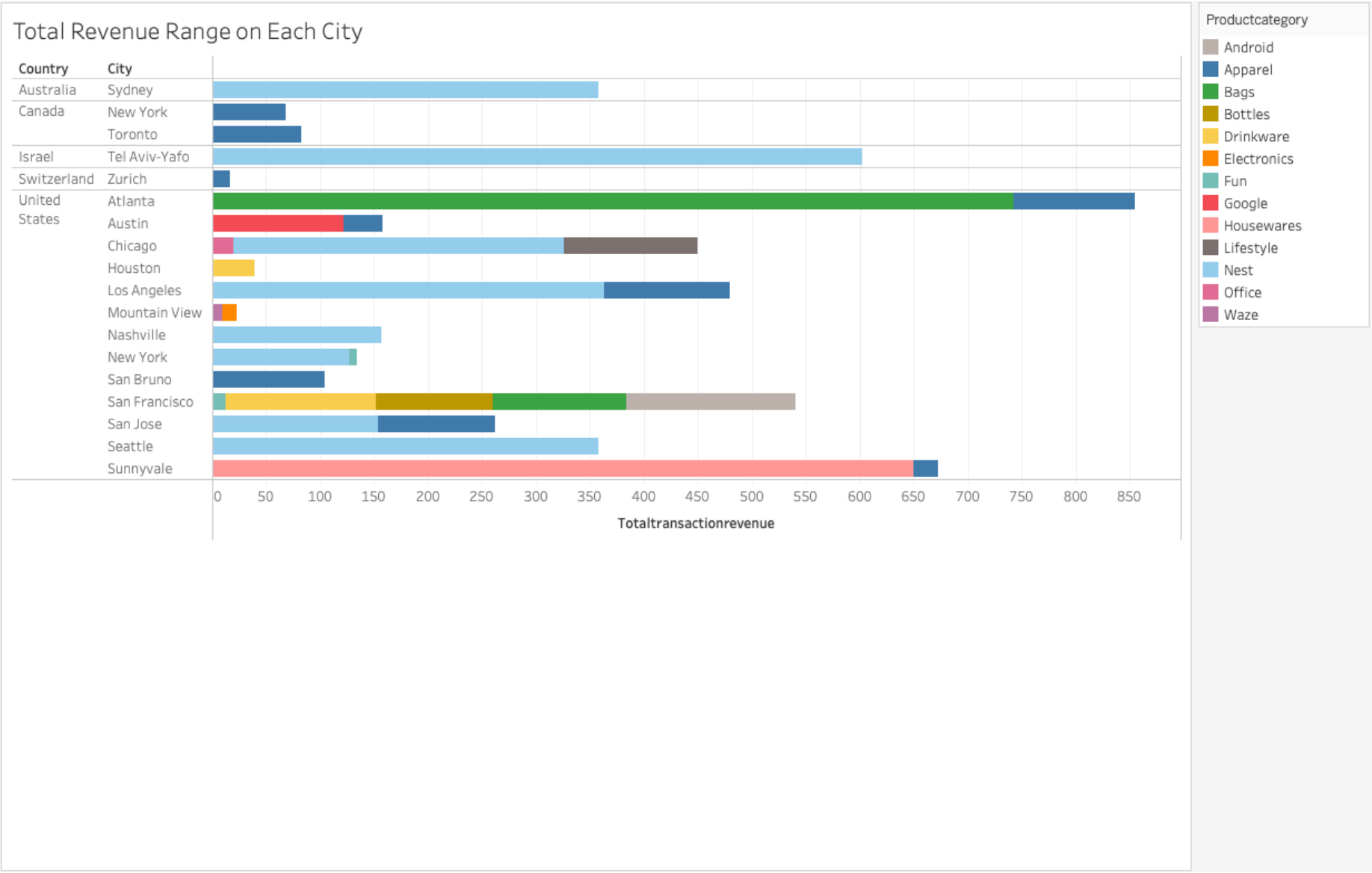


Productcategory

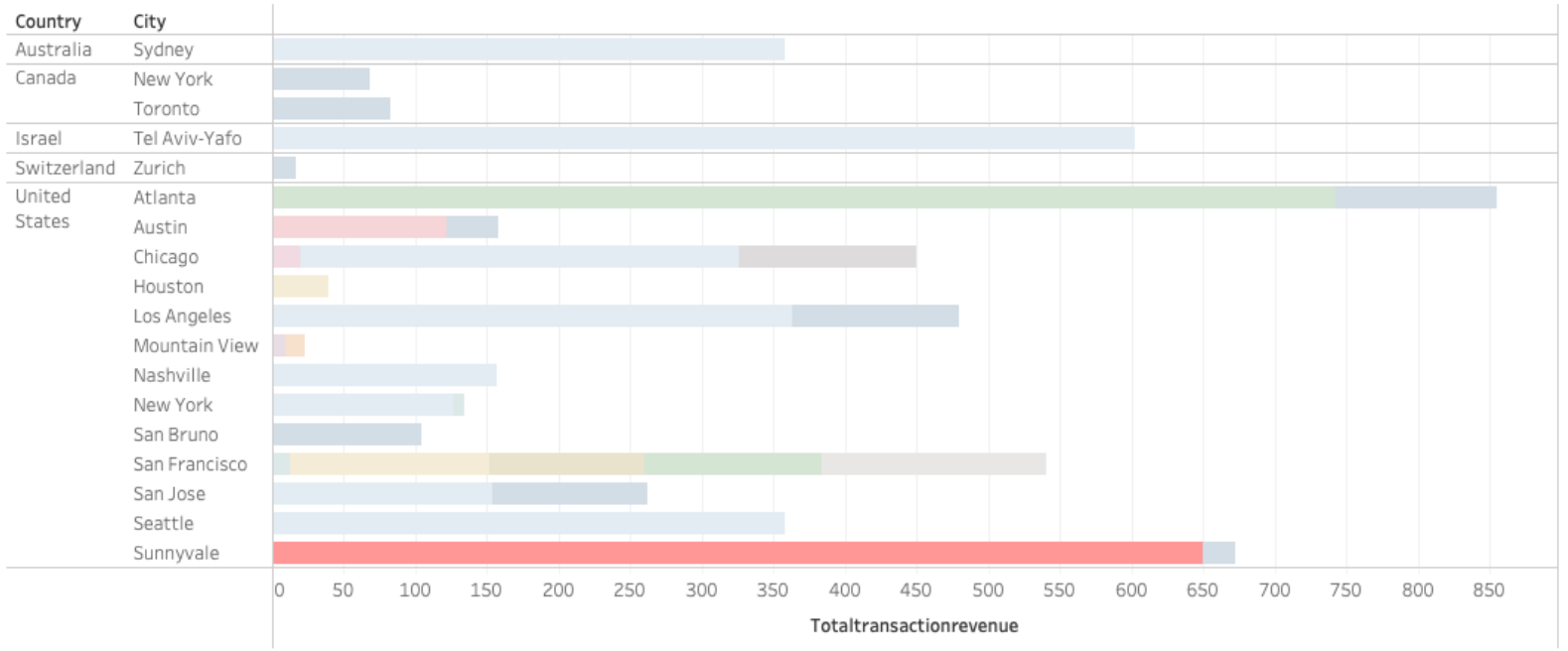
- Null
- Android
- Apparel
- Bags
- Bottles
- Drinkware
- Electronics
- Fun
- Google
- Lifestyle
- Nest
- Office
- Stickers
- YouTube

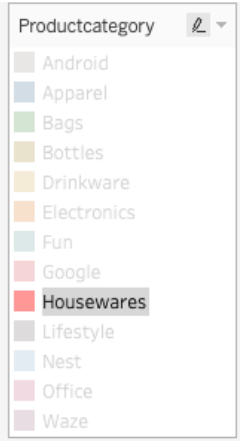
5. Can we summarize the impact of revenue generated from each city/country?

Total Revenue Range on Each City

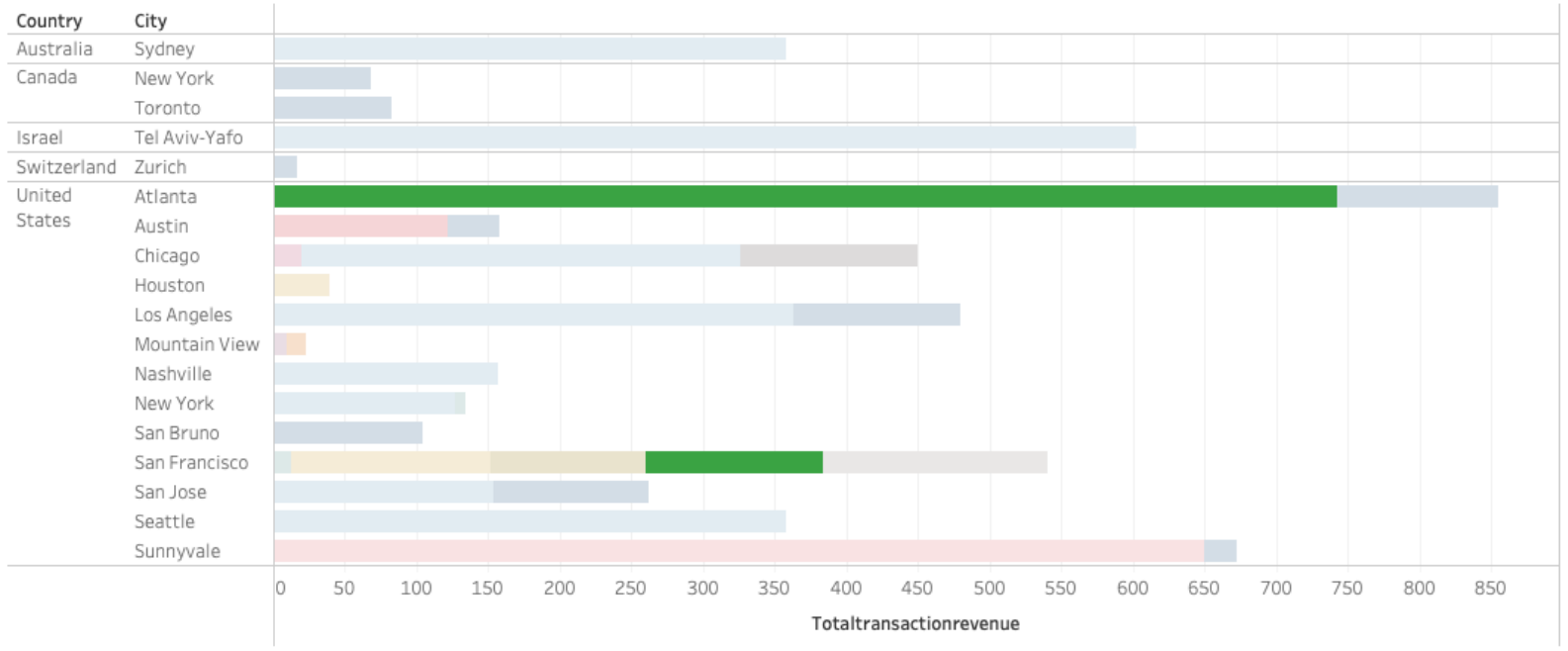


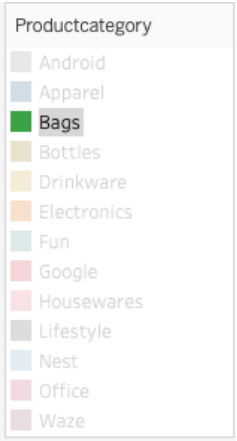
Total Revenue Range on Each City



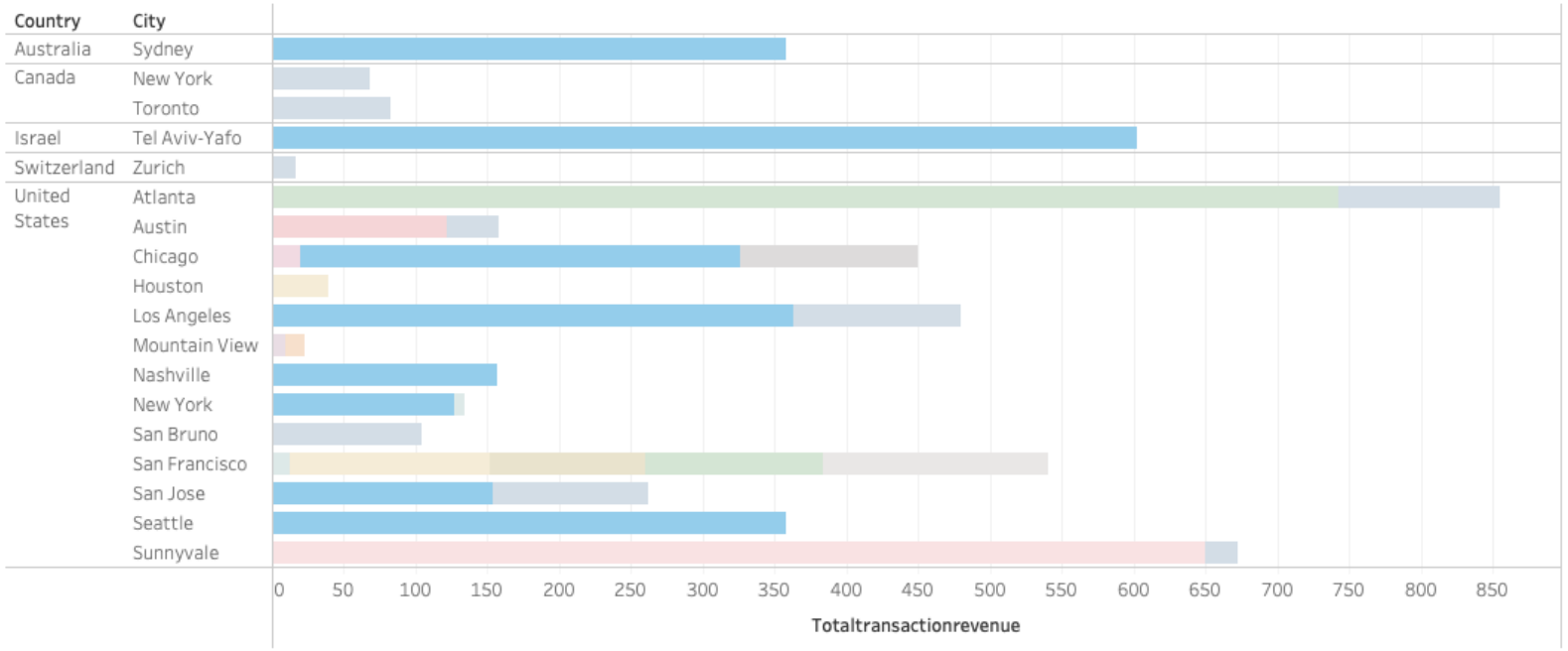


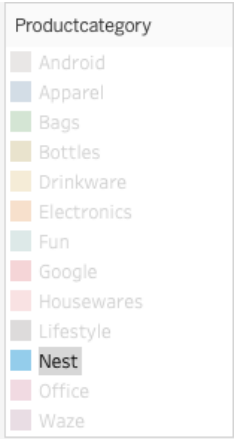
Total Revenue Range on Each City



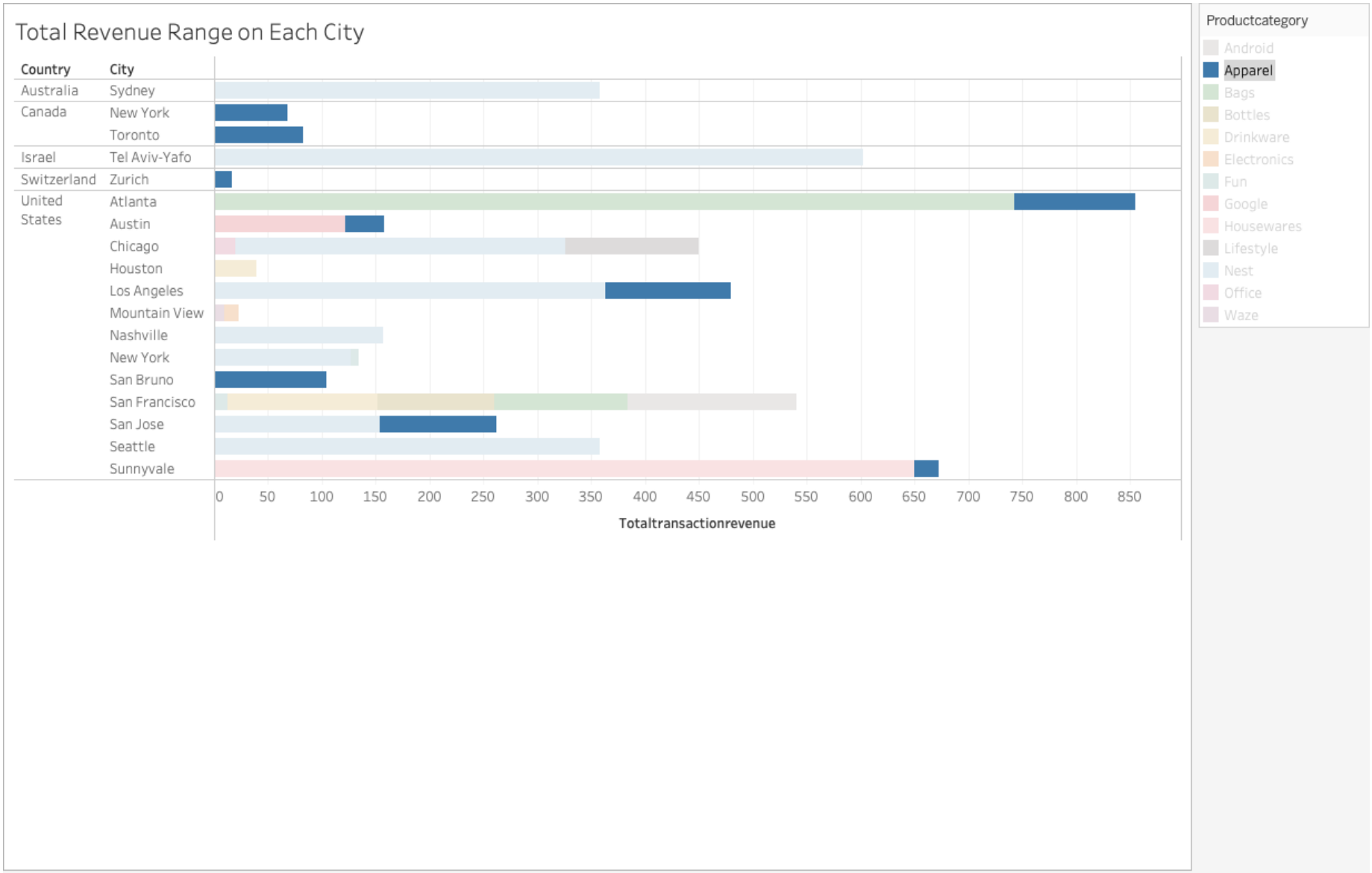


Total Revenue Range on Each City





Total Revenue Range on Each City



If I had more time

Recommendations

1. Proceed to the next data cleaning stage and data exploratory analysis; and
2. Conduct Natural Language Analysis on each of the top product category to determine the best product recommendations for each city/country.
3. Recommendation System