QVC Fulfillment-Does Speed Matter in E-Commerce?

Presented by: Evan Canfield, Khabirul Gainey, Joshua Ganz, Jake Hoertt, Julian Mucha, Harika Mudigonda

Agenda/Outline

- Background
- Data Processing
- Exploratory Data Analysis
- Questions
    - 1. Does the current distribution network maximize customer penetration (spend)? If not, what
    - should QVC do to increase customer penetration with the current distribution network?
    - 2. Are there specific products or product categories that should be located in specific
    - distribution centers?
    - 3. Do customers that receive their product sooner purchase more than customers with longer
    - delivery times?
- Results Summary
- Conclusion
    - Improvements for next time
    - Takeaways/ questions

Background

Our goal is to examine QVC's customer geography, distribution network, product mix, and purchase patterns to help QVC understand the relationship between speed of product/package delivery and customer loyalty.

Data

- QVC Sales Data
  - Three large spreadsheets (named QVC Data 1, 2, 3), each with just under 1 million rows
- Distribution Center Data
  - Warehouse ID, City, State, Cip Code
- Order Status Type
- Data Dictionary

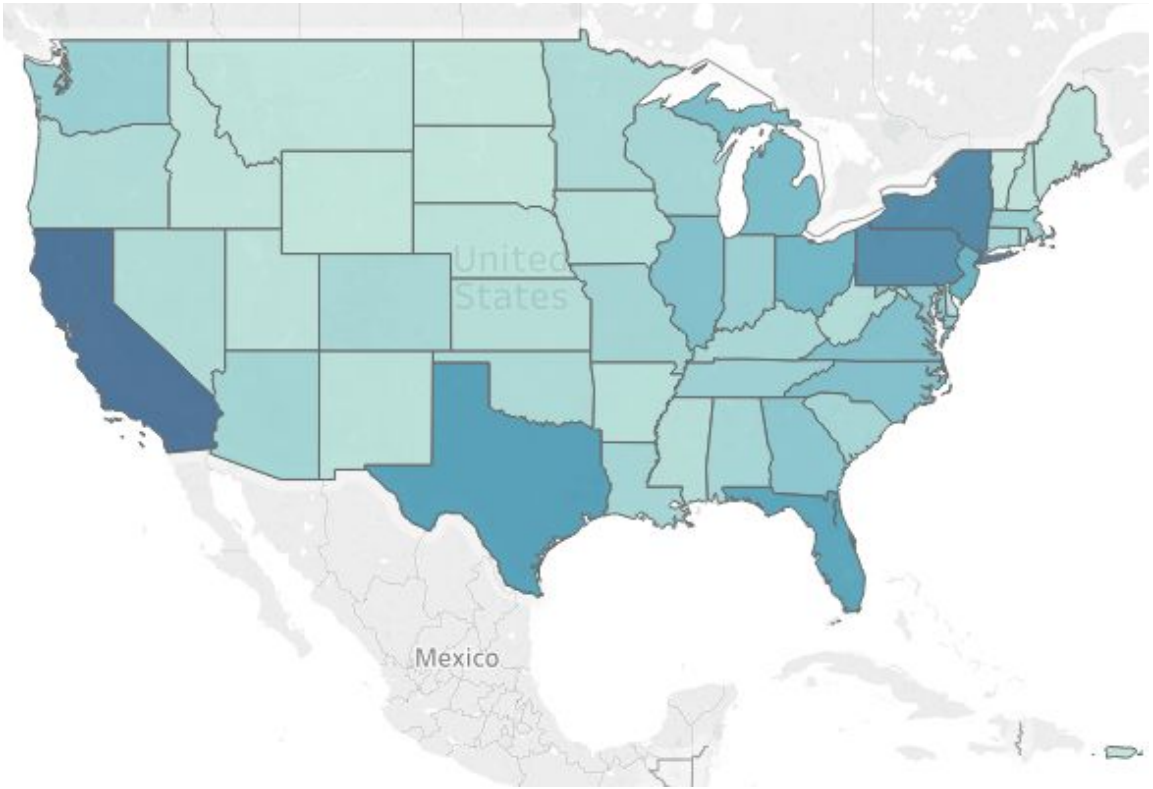
Derived Variables

- Rescheduled
  - Binary based on Rescheduled Date
- Distance
  - Haversine method via latitude and longitude of origin and destination
- Fulfillment Time
  - From Order date to Delivery date

Does the current distribution network maximize customer penetration (spend)? If not, what should QVC do to increase customer penetration with the current distribution network?

Question 1

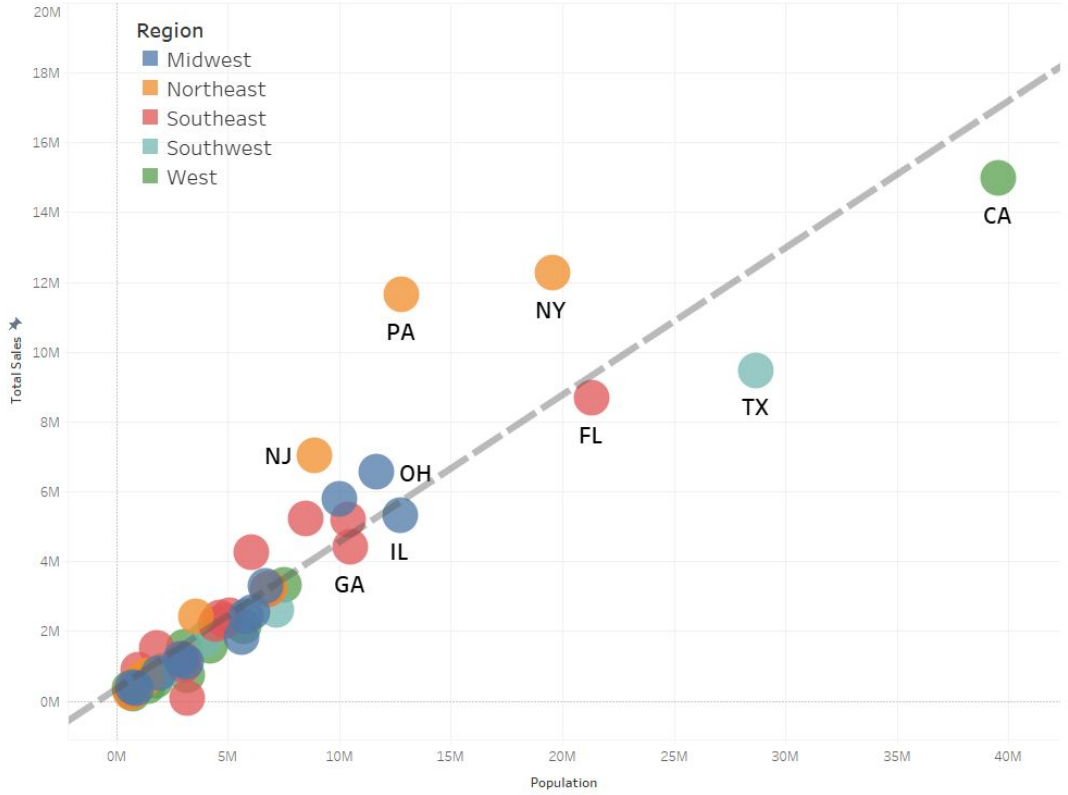
Sales By State



Top 5 States By Sales

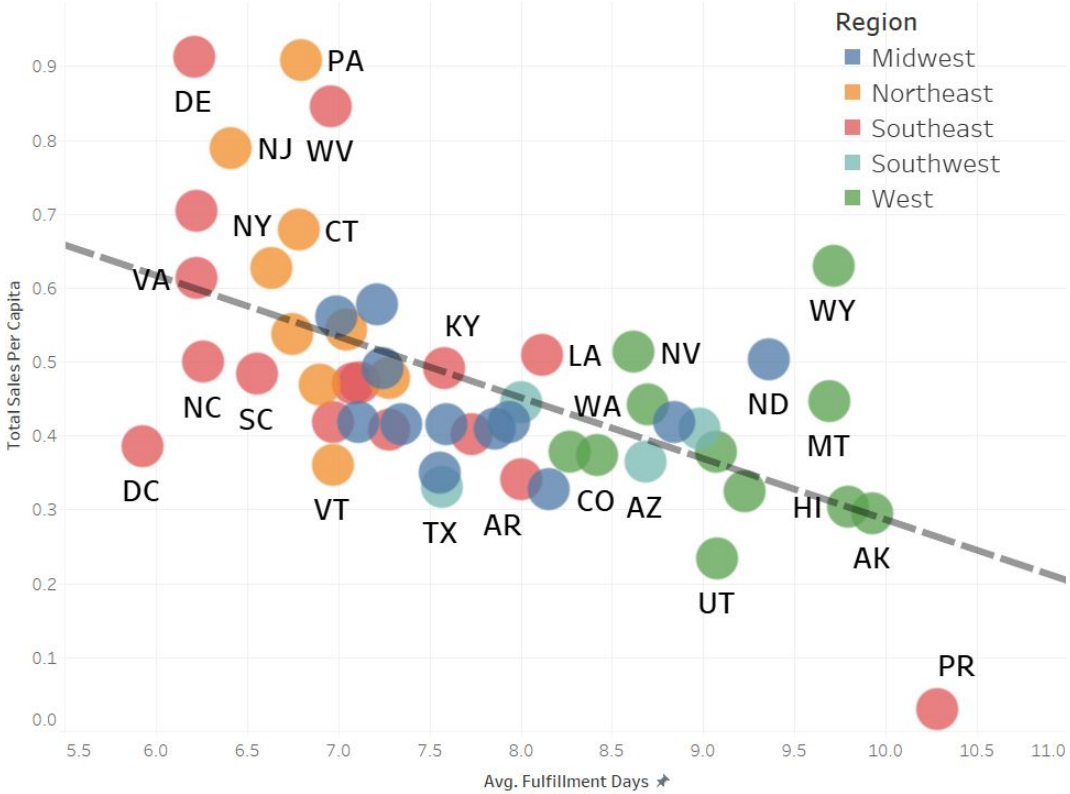
1. California
2. New York
3. Pennsylvania
4. Texas
5. Florida

Total Sales vs State Population



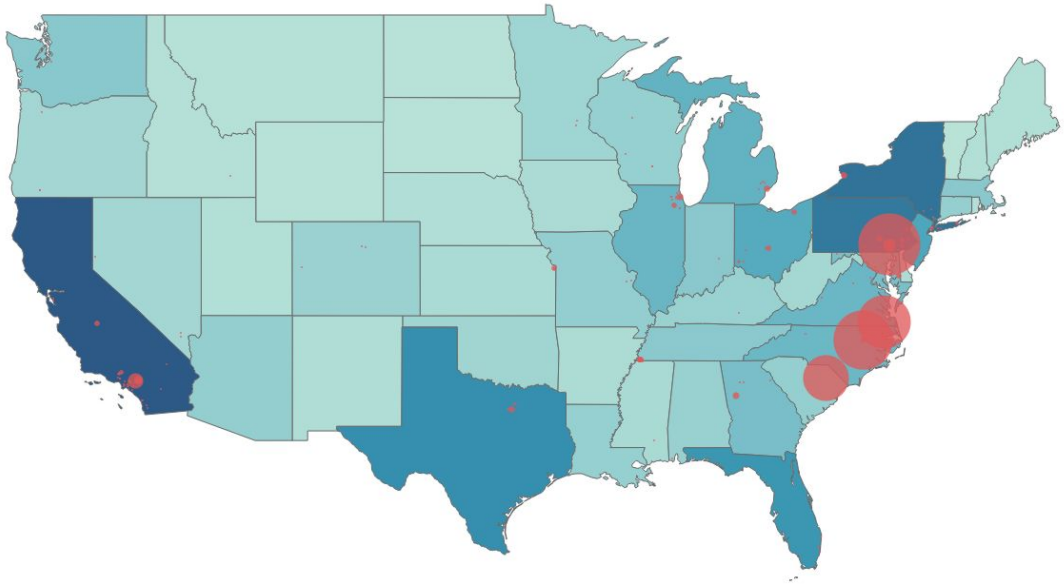
Total Sales vs Baseline Expectation

Sales per Capita vs Fulfillment Days



Shipping Warehouse Locations

- 91.2% of all sales are shipped from four Distribution Centers.
  - Located in PA, NC , SC, VA
- Largest Distribution Center in Texas handles 0.31% of all sales in the country
- Largest Distribution Center in California handles 1.7% of all sales in the country

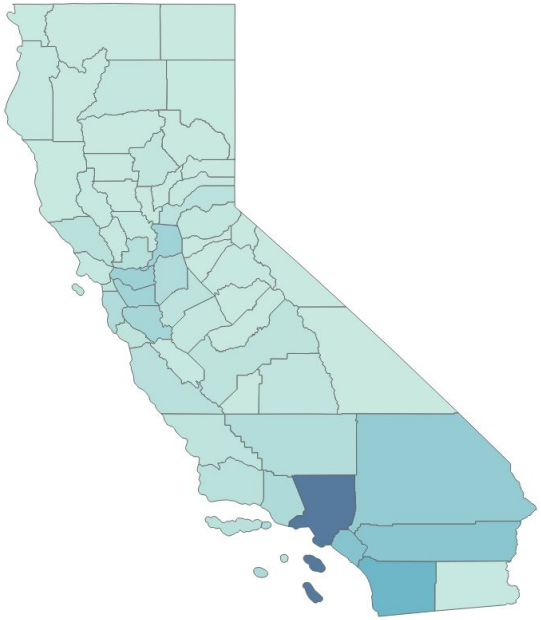


Question 1 Breakdown

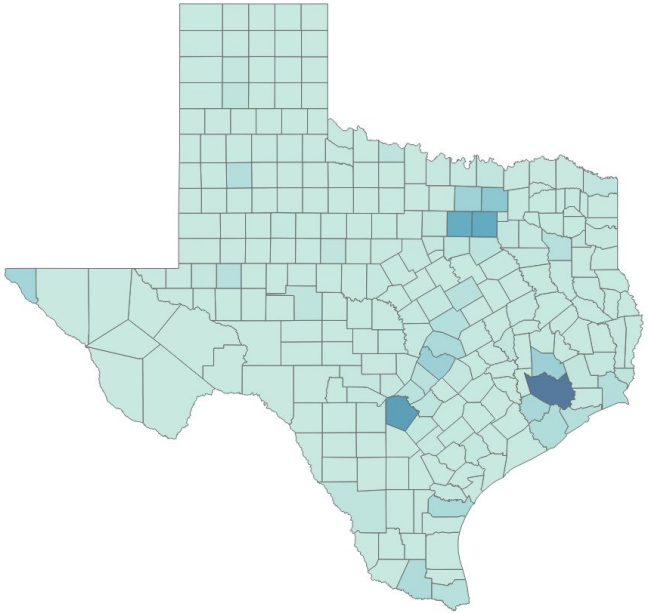
1. The State Sales - Population is linear, and provides the baseline performance expectation for each state
2. Some states over perform (Pennsylvania, New York) while other states under perform (Texas, California)
3. Under performance is correlated with longer order fulfillment times
4. Currently 91.2% of all shipping capacity comes out of four warehouses, all on the East Coast
5. **Recommend:** Expand or construct new warehouses in Texas, California, and Florida to reduce shipping times to these and surrounding states.

Possible Distribution Center Locations

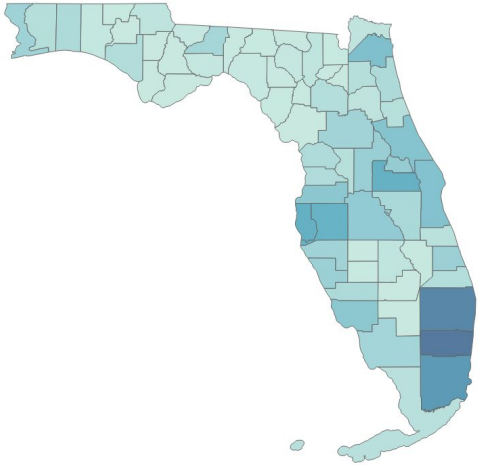
California



Texas



Florida

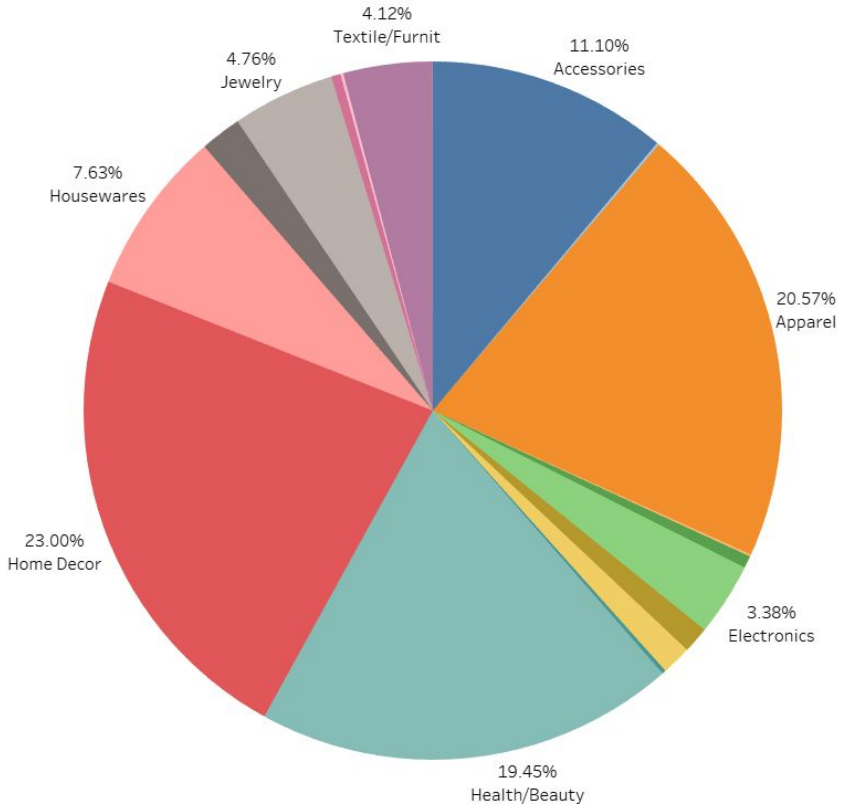


Are there specific products or product categories that should be located in specific distribution centers?

Question 2

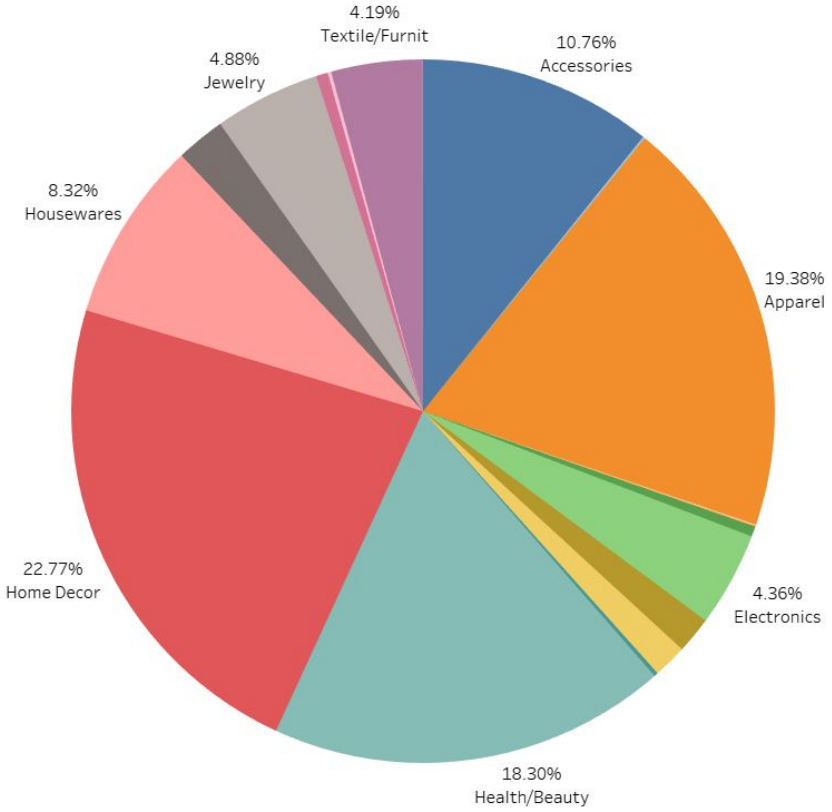
Product Categories Based on Products Shipped to CA

Top Product Categories	
Apparel	20.57%
Health/Beauty	19.45%
Home Decor	23.00%
Housewares	7.63%
Total	70.65%



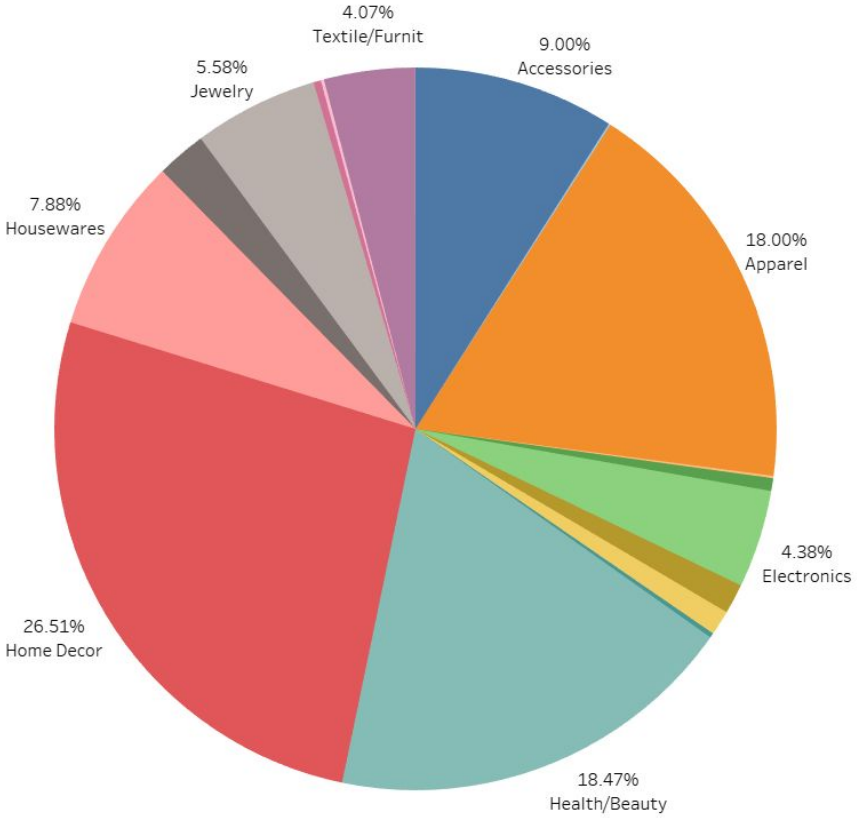
Product Categories Based on Products Shipped to TX

Top Product Categories	
Apparel	19.38%
Health/Beauty	18.30%
Home Decor	22.77%
Housewares	8.32%
Total	68.77%



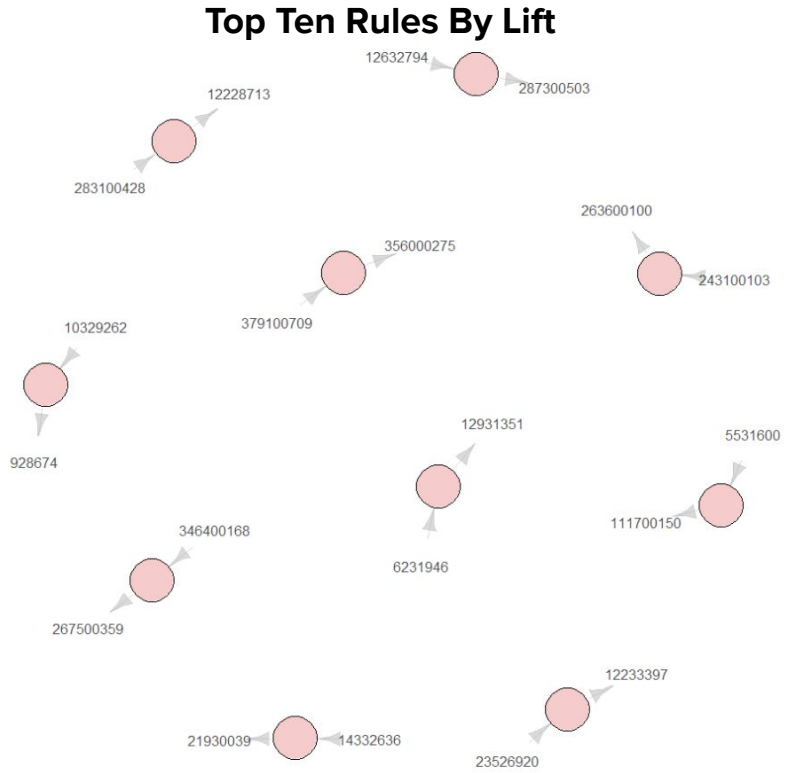
Product Categories Based on Products Shipped to FL

Top Product Categories	
Apparel	18.00%
Health/Beauty	18.47%
Home Decor	26.51%
Housewares	7.88%
Total	70.86%



Association Rule Mining

- Association Rule Mining revealed the following :
  - 611 unique rules
  - Five frequent sets of two
- Very low support threshold required to generate any rules or sets
- All rules were based on unique pairs



Question 2 Breakdown

1. Roughly 70% of sales in each state fall under the same four categories
    a. Apparel
    b. Health/Beauty
    c. Home Decor
    d. Housewares
2. Stock local distribution centers
    a. Majority of products from the four mentioned categories
    b. Goal of decreasing fulfillment time

Do customers that receive their product sooner purchase more than customers with longer delivery times?

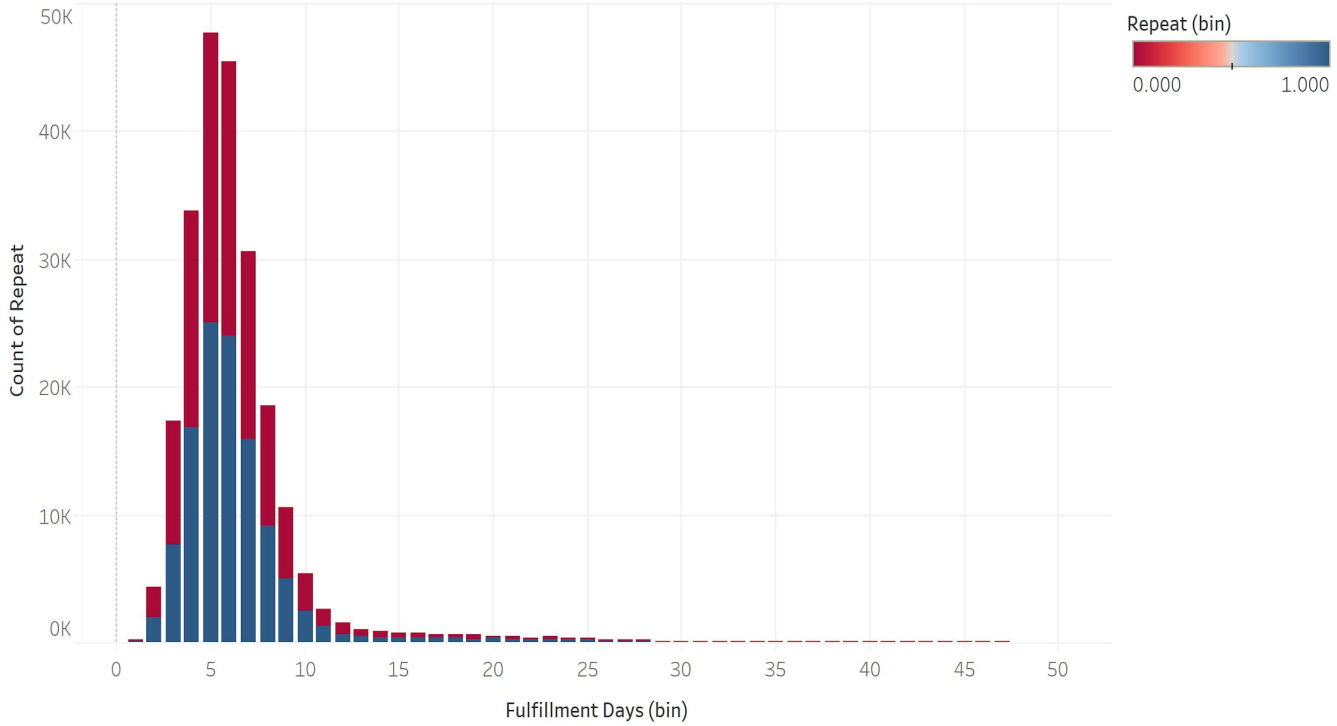
Question 3

Target Variable Creation

Create Repeat Customer labels	Filter down to first-order	Final Dataset with targets
Perform group by Party ID and count number of orders. Tag customers who made more than 1 order as repeat customers.	To ensure a fair model evaluation of identifying repeat customers, we must only use customers' first orders to train our model.	Our result is a customer level dataset with each row representing their first order, and a binary label indicating if they are a repeat customer.

Repeat Customer?

Repeat Customers Based on Fulfillment Days



Model Approach

Logistic Regression

- Examine confusion matrix
- Identify if there is statistical significance with variables related to order fulfillment time

Random Forest

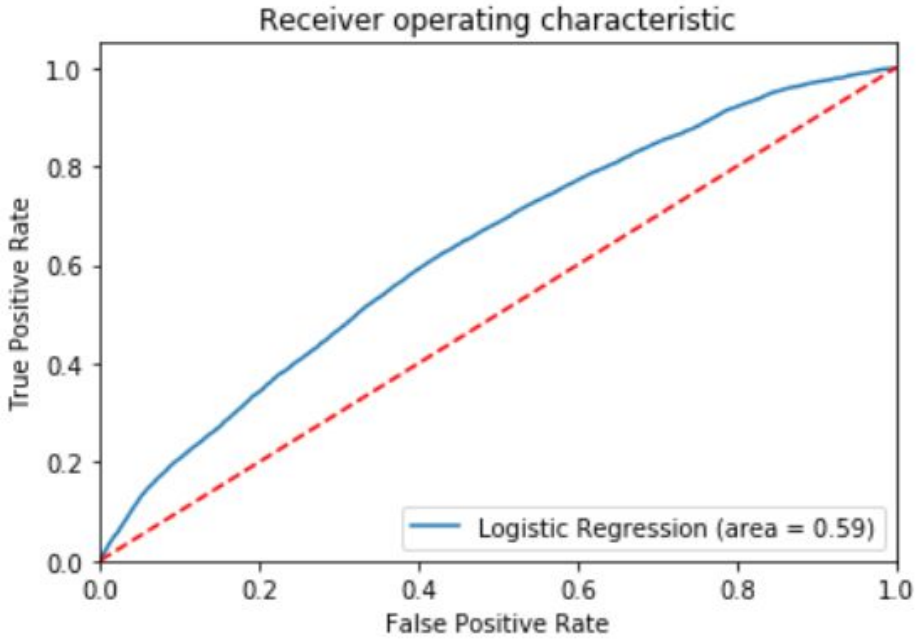
- Examine confusion matrix
- Identify influential variables in variable importance plots
- Less interpretable results, but more insight into feature set

Logistic Regression

```
Accuracy of logistic regression classifier on test set: 0.59
[[14235 14417]
 [ 9172 20302]]
              precision    recall  f1-score   support

           0       0.61      0.50      0.55     28652
           1       0.58      0.69      0.63     29474

   micro avg       0.59      0.59      0.59     58126
   macro avg       0.60      0.59      0.59     58126
weighted avg       0.60      0.59      0.59     58126
```

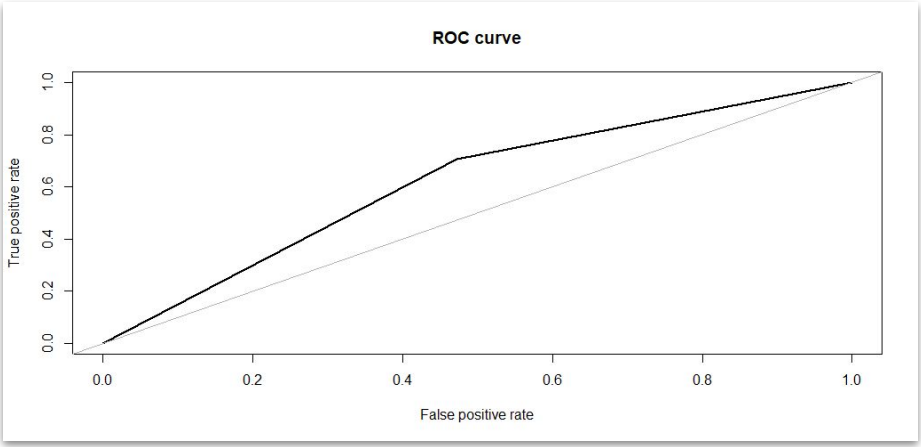


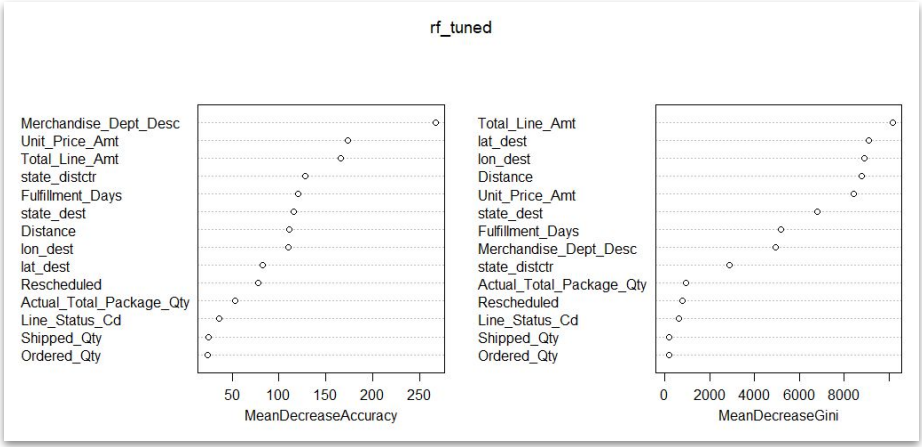
Random Forest

Tuned Model

- 1000 trees
- 3 variables selected for each tree
- AUC: 0.616

Fulfillment Days is a top 7 variable in both variable importance plots, but we observe that total price paid by customer is ranked higher in each plot.





Question 3 Breakdown

1. Unit price paid and Total price paid are more impactful than fulfillment days in determining repeat customers.
2. Having to reschedule an order made a customer less likely to be a repeat customer.
3. Distance and fulfillment days are related to each other.

Results Summary

Question 1: Does the current distribution network maximize customer penetration (spend)? If not, what should QVC do to increase customer penetration with the current distribution network?

- The current distribution network does not maximize customer penetration
- QVC should build, or expand existing, distribution centers in California, Texas, and Florida to decrease order fulfillment times in these customer dense areas

Question 2: Are there specific products or product categories that should be located in specific distribution centers?

- Local distribution centers in CA, TX, and FL should stock items in the apparel, health/beauty, home decor, and housewares categories since that is about 70% of the sales in each state

Question 3: Do customers that receive their product sooner purchase more than customers with longer delivery times?

- The random forest suggests that fulfillment days is an important variable in predicting repeat customers, but price paid may play a larger role.

Conclusion

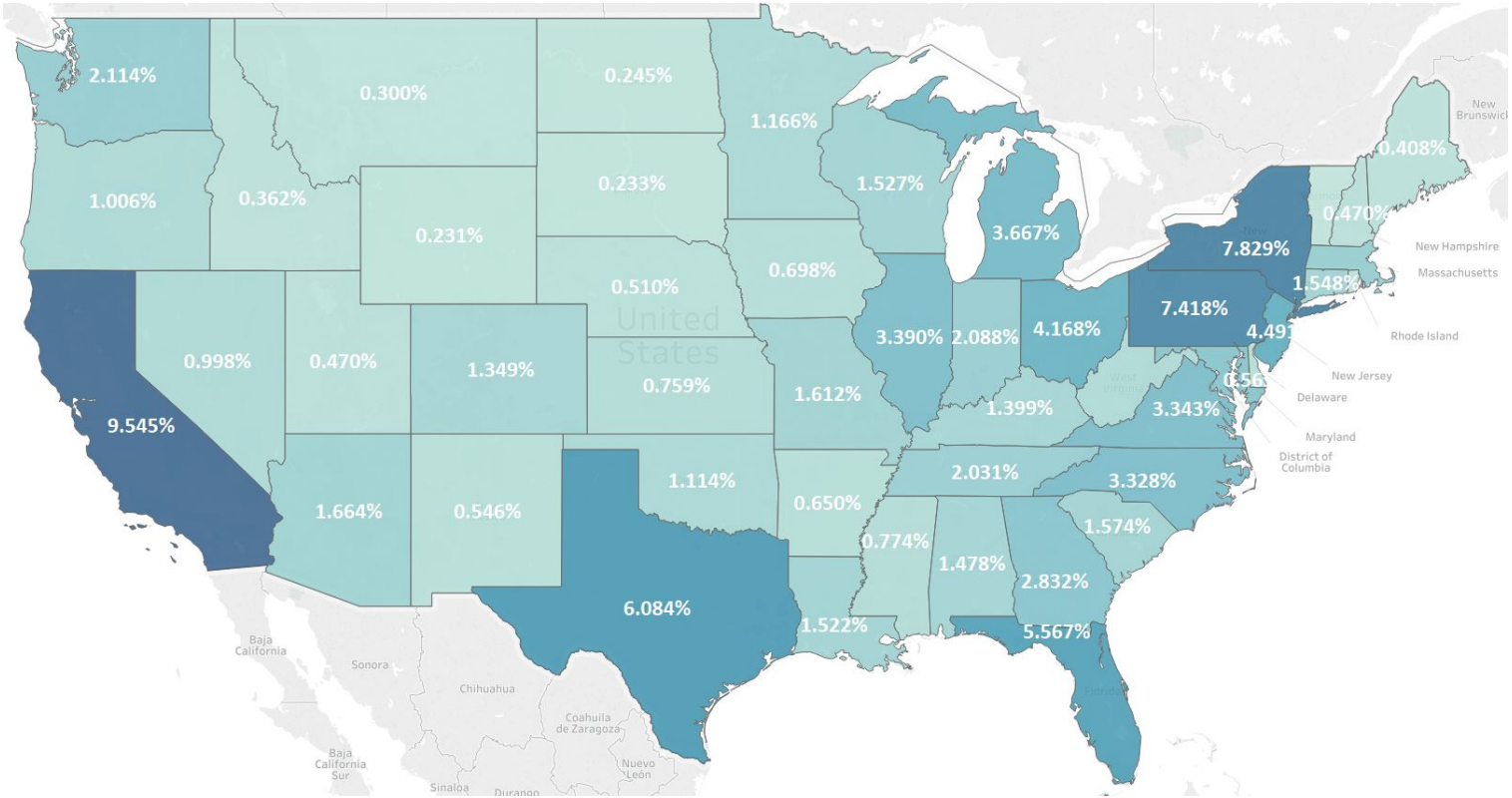
Improvements for Next Time

- More granular product data so we can provide more actionable insights regarding what products to stock at new and existing distribution centers.
- Connect our classification models predicting repeat business to a customer loyalty program to push targeted advertising.
- Better understand the relationship between company practices and data collection.
- Perform market research and competitive analysis on similar companies.

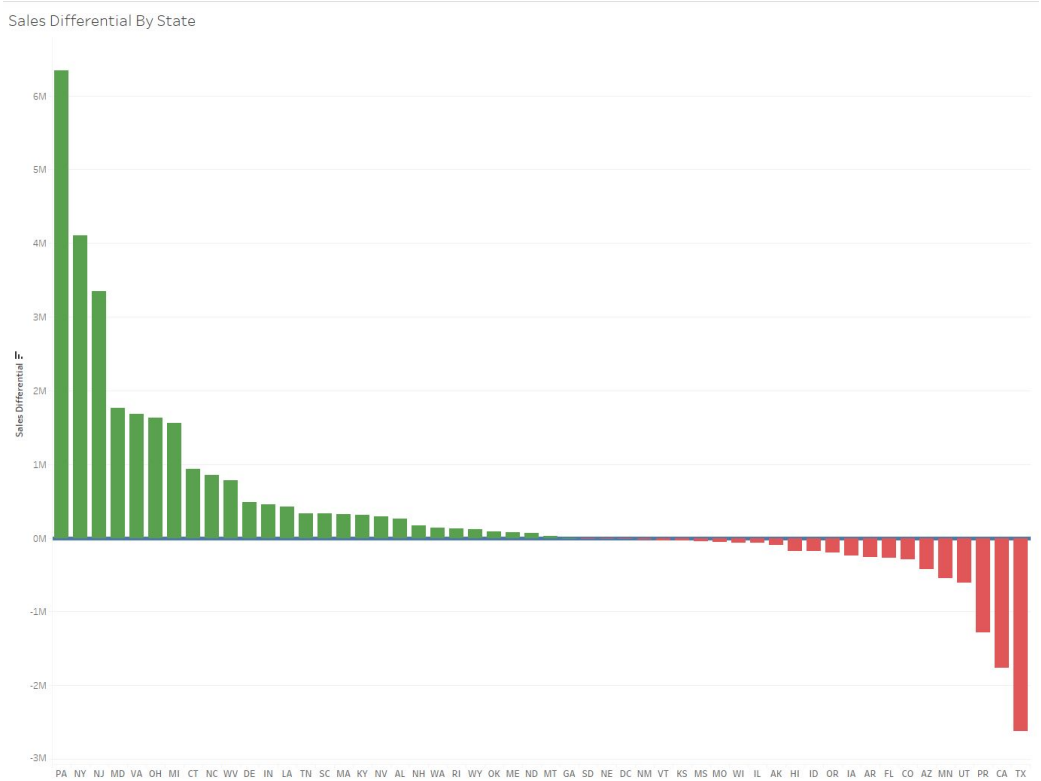
Questions?

Appendix

Sales By State



Total Sales vs Baseline Expectation



Association Rule Mining

Frequent Item Sets

items	support	count
{340000298,364705742}	3.59e-05	86
{281501189,314400011}	1.42e-05	34
{275300909,349600134}	1.09e-05	26
{284701163,357700000}	1.09e-05	26
{266000964,280100330}	1.00e-05	24

Association Mining Rules

lhs	rhs	support	confidence	lift	count
{10329262}	{928674}	4e-07	1	798029	1
{283100428}	{12228713}	4e-07	1	798029	1
{243100103}	{263600100}	4e-07	1	798029	1
{346400168}	{267500359}	4e-07	1	798029	1
{14332636}	{21930039}	4e-07	1	798029	1