# Variables

* From Order to Delivery (Days)
* Rescheduled? (Binary)
* Late (Binary)
* Days Late (Days)
* Distance – Distribution Center to Delivery Zip Codes

# Modeling / Analysis

* Correlation Matrix of Numeric
  + Particularly, what is the correlation between Distance, Late, Amount Spent? Number of items?
* Retention
  + Construct Data Frame around the customer
  + Possible Variables
    - Late Binary
    - Early Binary
    - Average Late Packages
    - Average number of days off the expected date
    - Dependent: Multiple Packages or No
* Rescheduled Correlations
  + Number of Packages
  + Distance
  + Department

# Questions

* What makes people more likely to re-order? Retention Rate!
  + On time delivery?
  + Short delivery times?
  + Large purchase amounts?
* If new distribution centers are opened, what items should be where? What is each state most interested in.

# Visualization Ideas

1. Map
   1. Start with State Choropleth map based on sales
      1. Can you overlay highways / railroads
   2. Then overlay Distribution Center Points
2. Line graph - Distance vs Estimated Deliver Times
3. Line graph – Sales vs. Population
   1. This will show penetration
4. By State / Region - Percent of deliveries
   1. Out of State
   2. Farther than x miles (100?)
5. County Choropleth
   1. For states where we would recommend a distribution center. Where would it be?
   2. Convert Sales Zip Code to county / FIPS
   3. <https://community.tableau.com/thread/289855>
6. Heat Map for Departments

# Project 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?
   1. Spending vs Population
      1. Point out states below the line
   2. Spend per Population vs Average Distance or Delivery Times
      1. I expect state with Longer Distance / Delivery Times to have lower spend per pop
   3. Install distribution centers in certain areas
      1. California, Florida, Texas
      2. Show County Choropleths for select states/regions
         1. Inspect Omaha, NE /Kansas City / St. Louis /Chicago but look the region, not just the state
2. Are there specific products or product categories that should be located in specific distribution centers?
   1. Determine by state what categories they spend on. Adjust Distribution Centers by nearby states
3. Do customers that receive their product sooner purchase more than customers with longer delivery times?
   1. Create customer data frame
   2. Create dependent variable
      1. Customer associate with more than one purchase - Binary
   3. Logistic regression – inspect early package variable coefficient for effect

Notes

* Choropleth showing percent sales by state
* Choropleth showing difference to Sales per Capital Average (slope of regression line)