**E-Mail Analytics dataset**

**Background:**

This dataset contains 64,000 customers who were involved in an e-mail test.

1/3 were randomly chosen to receive an e-mail campaign featuring Mens merchandise.

1/3 were randomly chosen to receive an e-mail campaign featuring Womens merchandise.

1/3 were randomly chosen to not receive an e-mail campaign.

During a period of two weeks following the e-mail campaign, results were tracked. Your job is find if the Mens or Womens e-mail campaign was successful.

**Data Attributes:**

Recency: Months since last purchase.

History\_Segment: Categorization of dollars spent in the past year.

History: Actual dollar value spent in the past year.

Mens: 1/0 indicator, 1 = customer purchased Mens merchandise in the past year.

Womens: 1/0 indicator, 1 = customer purchased Womens merchandise in the past year.

Zip\_Code: Classifies zip code as Urban, Suburban, or Rural.

Newbie: 1/0 indicator, 1 = New customer in the past twelve months.

Channel: Describes the channels the customer purchased from in the past year.

Another variable describes the e-mail campaign the customer received:

Segment

Mens E-Mail

Womens E-Mail

No E-Mail

Finally, we have a series of variables describing activity in the two weeks following delivery of the e-mail campaign:

Visit: 1/0 indicator, 1 = Customer visited website in the following two weeks.

Conversion: 1/0 indicator, 1 = Customer purchased merchandise in the following two weeks.

Spend: Actual dollars spent in the following two weeks.

**Required questions:**

1. Use proc contents to get familiar with the structure of the data
2. Check the distribution of the data by email type (Segment)
3. Find the statistical properties of visit, conversion, and spend.
4. Find the statistical properties of visit, conversion, and spend by email type (Segment). Which e-mail campaign performed the best, the Mens version, or the Womens version?

**Optional questions (if you have time… you do not need to solve them all):**

1. How much incremental sales per customer did the Mens version of the e-mail campaign drive? How much incremental sales per customer did the Womens version of the e-mail campaign drive?
2. If you could only send an e-mail campaign to the best 10,000 customers, which customers would receive the e-mail campaign? Why?
3. If you had to eliminate 10,000 customers from receiving an e-mail campaign, which customers would you suppress from the campaign? Why?
4. Did the Mens version of the e-mail campaign perform different than the Womens version of the e-mail campaign, across various customer segments?
5. Did the campaigns perform different when measured across different metrics, like Visitors, Conversion, and Total Spend?
6. Which audience would you target the Mens version to, and the Womens version to, given the results of the test? What data do you have to support your recommendation?

**Reference:**

<http://blog.minethatdata.com/2008/03/minethatdata-e-mail-analytics-and-data.html>