

**UNIVERSITY OF NORTH CAROLINA AT CHARLOTTE**

**ADVANCED BUSINESS INTELLIGENCE AND ANALYTICS**

**Better Product Sales for QVC**

**Analyzing Airtime impact on product sales with different demographics and brands**



**About QVC**

QVC is an acronym for "Quality, Value, Convenience" is an American cable, satellite and broadcast television network, and multinational corporation specializing in televised home shopping. QVC was founded on June 13, 1986 by Joseph Segel. All of QVC's operations (U.S., UK, Germany, Japan, Italy, France, and China) run 24 hours a day, although live programming hours vary between each region.

Today, QVC is the world's leading video and ecommerce retailer, offering a curated collection of desirable brands to millions of customers around the globe each day through broadcast, Internet, and mobile sales platforms. QVC is the world's top multimedia shopping company in terms of viewers and revenue and ranks among the top online mass merchandise retailers. QVC's 2014 sales generated $8.8B in annual revenues, $3.5B of which came from eCommerce.



In this process, we want to understand correlation between the product’s airtime and buying patterns from customer with couple of other questions related to customers and product behavior.

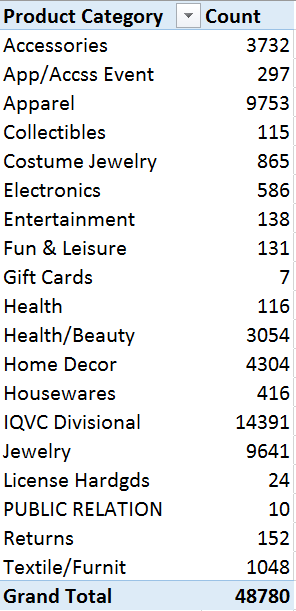
**SUMMARY STATISTICS**

Some of the important data that we have extracted from various QVC data sets provided to us.

**Summary statistics of Product Master:**

The product master consists of Product Nbr, Product description, Product category and Brand name.

* Analyzed the total number of products in each product category
* The total number of products is 48,780
* The total number of product categories is 19
* The IQVC Divisional has highest number of products - 14,391 followed by the Apparel - 9,753
* The least number of products sold is of the Gift category - 7



We also did a pivot table to find out which brand had the highest sold products

* Approximately there are total 6,367 null values
* The top brand is Susan Graver, 1814 products sold
* There is a total of 1884 brands in the QVC



**Summary Statistics of Product Airtime**

* Total Product Categories considered for product airtime – 16
* Total Brands under the categories for airtime – 654
* Product Category which had the highest airtime – Home Décor – 872 Hours
* Brand with highest airtime – 146 Hours

**Air time for Product Categories**

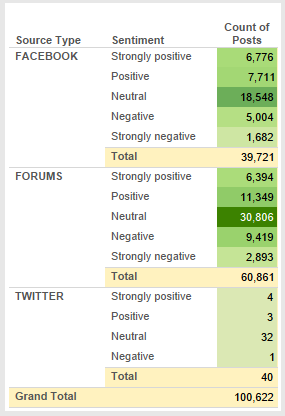
|  |  |
| --- | --- |
| **Product Categories** | **Air Time in Minutes** |
| Home Decor | 52372.77 |
| Apparel | 43579.81 |
| Accessories | 34943.45 |
| Jewelry | 29463.25 |
| Health/Beauty | 29174.26 |
| Housewares | 18561.15 |
| Electronics | 12341.72 |
| Textile/Furnit | 7199.02 |
| Fun & Leisure | 6760.99 |
| Costume Jewelry | 1834.68 |
| App/Accss Event | 1695.94 |
| Health | 1258.59 |
| Entertainment | 766.84 |
| IQVC Divisional | 578.83 |
| Collectibles | 171.49 |
| License Hardgds | 106.11 |
| **Grand Total** | **240808.9** |

**Air time for Top 20 Brands**

|  |  |
| --- | --- |
| **Brands** | **Air Time in Minutes** |
| Not Known | 52948.43 |
| Susan Graver | 8806.12 |
| Denim & Co. | 7524.06 |
| N/A | 5780.77 |
| Dooney & Bourke | 5419.29 |
| Isaac Mizrahi Live! | 5161.4 |
| Clarks | 5059.62 |
| Cottage Farms | 4693.72 |
| Roberta's | 4055.47 |
| IT Cosmetics | 3784.03 |
| Vionic with Orthaheel Technology | 3754.03 |
| philosophy | 3261.07 |
| Joan Rivers | 2908.12 |
| Dyson | 2828.46 |
| LOGO by Lori Goldstein | 2602.36 |
| Liz Claiborne New York | 2537.43 |
| Josie Maran | 2456.83 |
| Steel by Design | 2319.5 |
| Quacker Factory | 2243.56 |
| Dennis Basso | 2142.16 |

**Summary statistics of Social media:**

This would compared to the overall sentiment of all posts posted in the same time frame.  A positive relationship is expected however text analysis will be need to be done to confirm this.  Clustering terms will show the types of posts QVC receives.  Forecast Studio may be used to plot the weekly revenue.

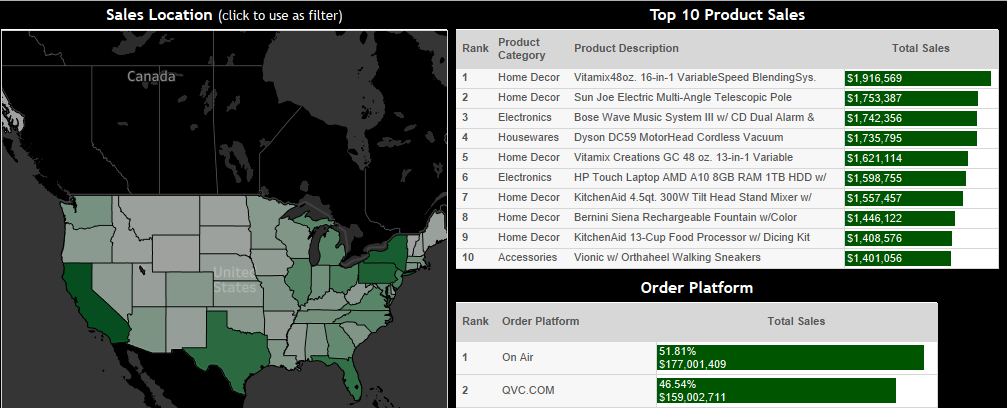


**We would like to analyze each question individually and provide recommendations for QVC to improve sales, based on our research.**

**Analysis**

**Question 1: What are the products and product categories that sell best in the US market?**

We focused on the “product\_master” and “order\_master” tables. Both tables were joined using product\_number column.

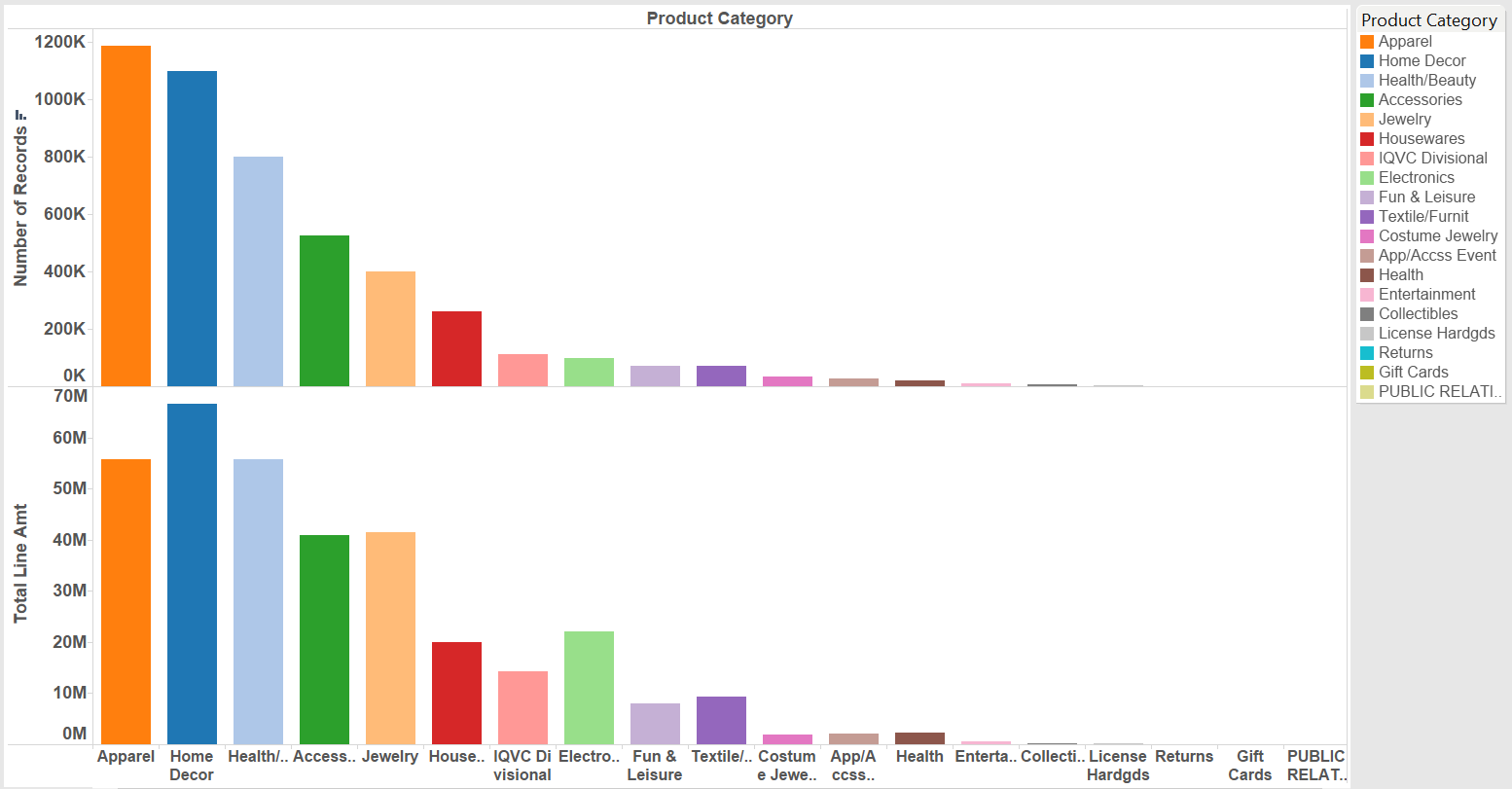


We developed an interactive Tableau dashboard [(see Tableau Public link here)](https://public.tableau.com/profile/bill.stuart#!/vizhome/QVCProject/QVCProject) that allows to dig deeper into popular sales items in the US market as well as at the state level. As you can see above, the most popular product is the **Vitamix Blender** and the most popular method of order is **on-air**.

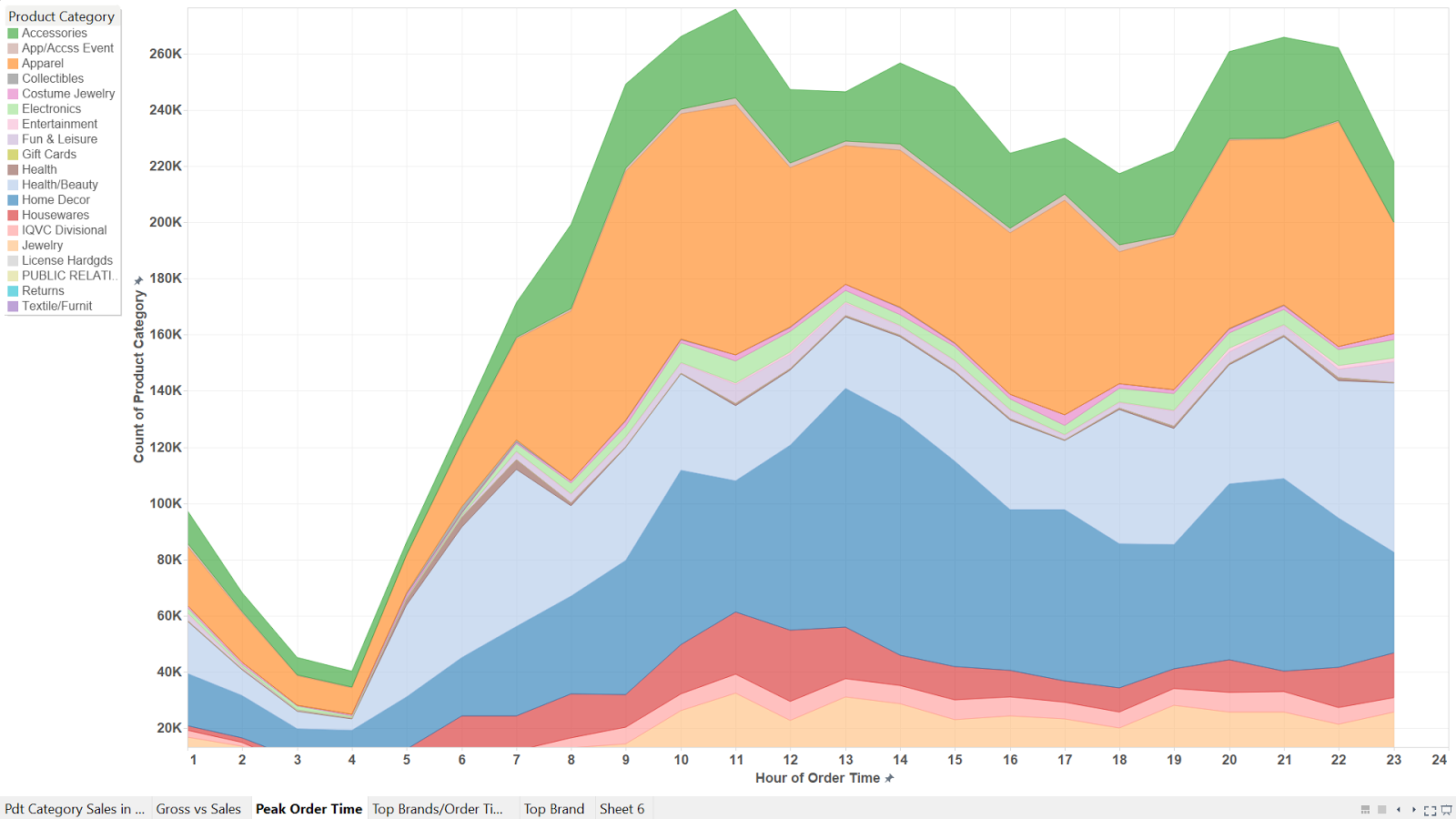
The dashboard acts as a filter as you click the state; as you dig down to the state level you can see that the top 10 products do not vary to an extreme, many are repeated. Further, you can see Home Décor accounts for 6 of the top 10, Electronics has 2, and Accessories and Houseware has 1 each.

This kind of visual analysis can be very effective, because you can easily scan states to see regional trends. **California** is the state with most total line amount. For example, we can see that in general on-air sales and qvc.com sales are about even. It may be important to see whether there are some states or regions where TV is more popular than the internet (less urban areas) and it’s more effective to choose email campaigns in this state. North Dakota, South Dakota, and Wyoming all have a 10-point difference in preferring TV to internet.

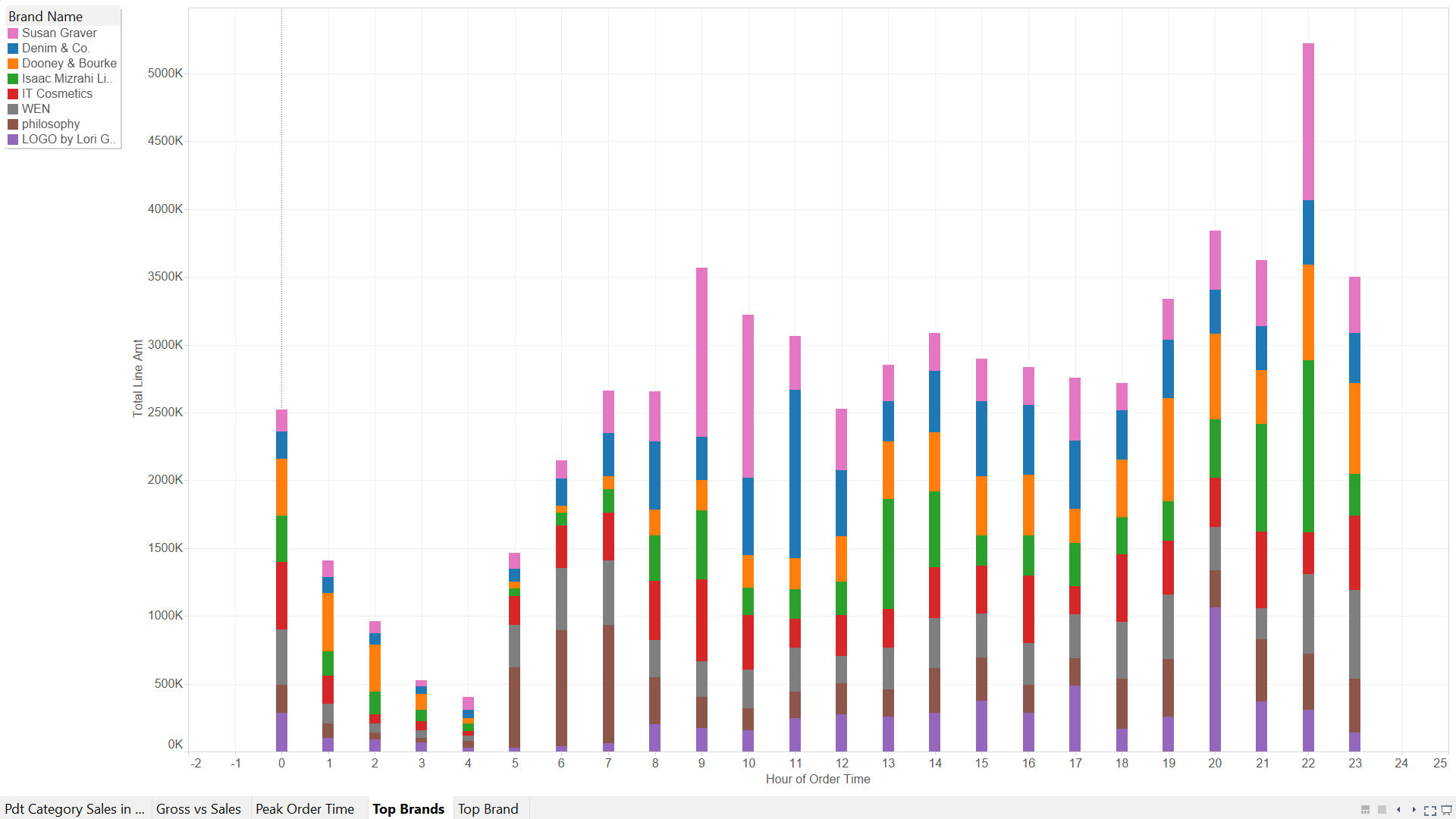
**Total Gross vs Total Sales**



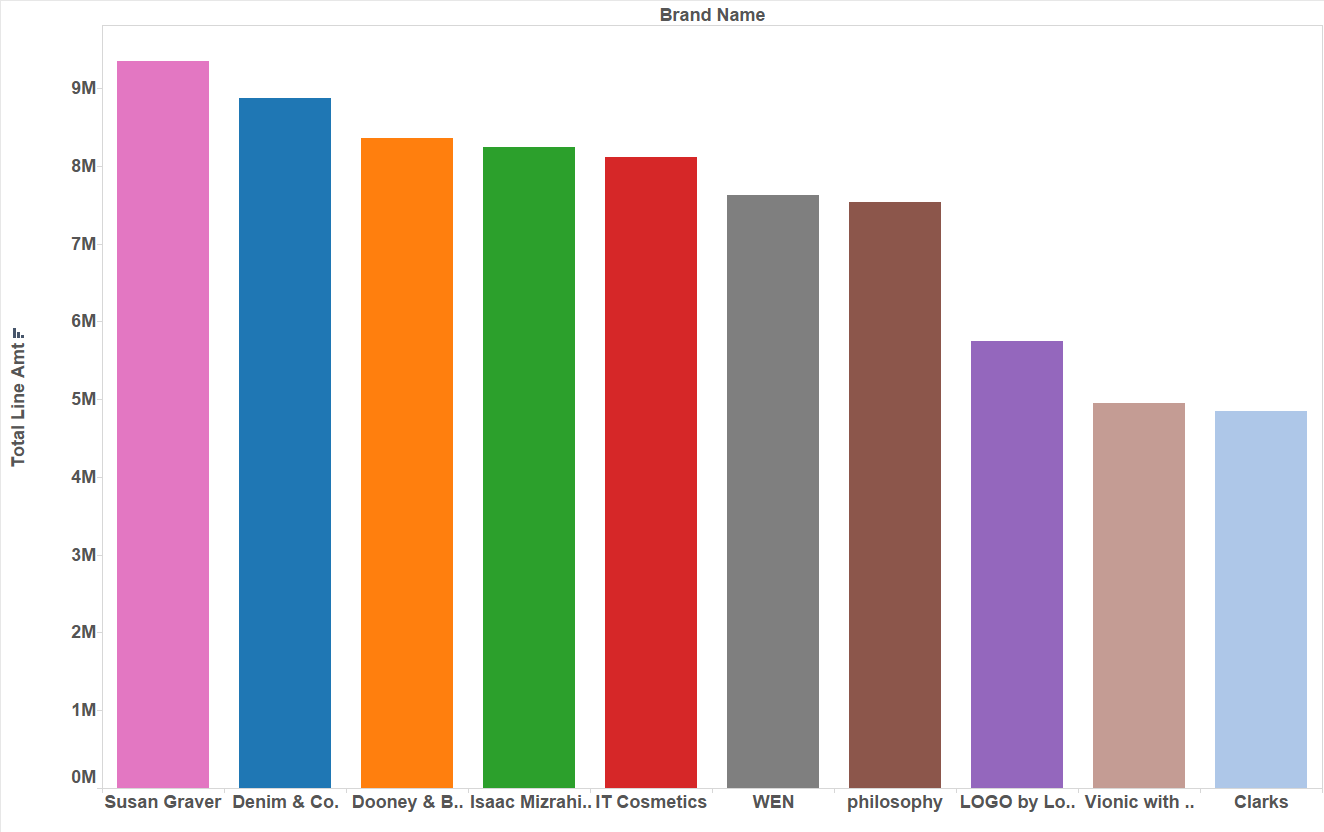
In this visualization, we are comparing the total number of products category sold versus the total amount of sales. More number of products are sold in the Apparel category but the total line amount is more for the Home décor category.  This means that average price of Home Décor products is more as compared to Apparel products.



This visualization depicts the order of time versus the product category. The peak time for all products is 11 AM in the morning, and accessories have the most number of sales with 280K. And once again in the evening the peak goes high at around 9 PM.



The graph shows which brands were sold during which time of the day. We can compare the above two graphs and also conclude that even though more products were sold at 11 am but more revenue was generated at night at 10 pm. So we can conclude that more revenue is generated by sales at night time than during day. The sales of Denim $ Co is more at 11 am for the 3 months.



This graph depicts the revenue generated by different Brands. **Susan Graver** is the most sold brand with revenue above 9M.   The second and the third highest would be Denim & co followed by Dooney.

**Question 2: How effective are QVC’s campaigns in driving product sales?**

The Campaigns done by QVC are through emails. The cost associated with each of the many email campaigns is given in the project data email\_campaign. We have categorized the campaign cost by product and month. We have clubbed together the different email campaigns done in each month into one set of campaigns. There have been 3 such months, namely, January, February and March. We have taken the start date of each month’s campaigns as the date of first such campaign of that particular month.

The first step to answer this question will be to join the product\_master and email\_campaign spreadsheet on product\_category and then join that resulting table with the order\_master spreadsheet on product\_NBR. In order to determine trends around when the email campaigns occurred the final spreadsheet will be analyzed in SAS Forecast Studio.  Using this program will allow us to see the significance level of the email campaign as well as determine if there are any seasonal or monthly trends. The before and after effect of the email campaign can also be seen.  Another potential process would be to use SAS Enterprise Miner (or Guide) and apply linear regression to the spreadsheet. If this is done revenue would be the dependent variable and the number of email campaigns per product would be the independent variable.  An assumption (an possible explorative result) would be that a product with a large number of email campaigns would have a large significant coefficient showing an increase in revenue.

In order to determine whether or not the email campaigns have been successful 2 separate models have been created. An inner join was done to join order master with product master, and that resulting table with the email campaign table. The first model uses the total line amount as the dependent variable and the amount of money spent for the email campaign as the independent variable.  The email campaign amount was grouped according to product categories. The total line amount was originally grouped by products, however that original model produced an R-Squared value of 0.0 so the model was scrapped. The total line amount was then grouped by order date, followed by order time.  A model was then created using the same variables in the same manner.  The resulting model is: Total\_Line\_Amount=90.633 +.00198campaign\_spend.  The standard error is very low, and although the R-Squared value is also very low this model is still usable.

Since the first model shows that the amount of money spent towards email campaigns does not make a significant effect on sales another model was created to see if the amount of money spent towards the email campaign would affect the number of products bought around the same time. Since total\_line\_amount isn’t necessary the two tables that were joined were product master and email campaign. The number of products in each product category was summed and grouped by product category. The email campaign amount was also summed by product category. The two tables were joined using an inner join.  The resulting model is: count\_of\_product\_category=5673.99 + .00002326campaign\_spend.  The model has a slightly better R-Squared value than the first one, and low standard errors. Overall the amount of money spent on an email campaign doesn’t affect either the sales or the number of products purchased. Therefore the email campaigns are not effective and should be either removed completely or explored more to see how the efficiency could be raised

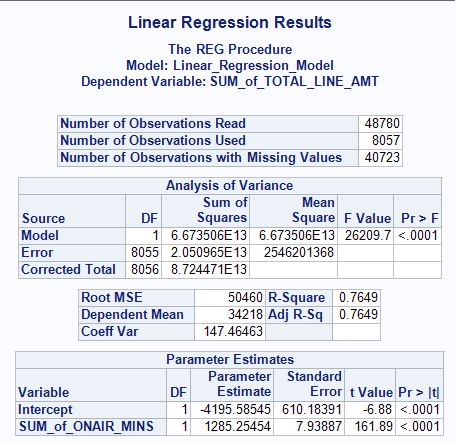
**Question 3: How effective are QVC’s on-air broadcast and on-air personalities in driving product sales?**

For this analysis we considered “order\_master” and “product\_airtime” tables. This was achieved by joining Order Date with Order Time and comparing with On Air Time.

One of our main objectives is to determine whether the on-air broadcast and on-air personalities drive sales. Since QVC’s advertising is through television, we would expect a very high correlation between these variables.

To prepare the data, we used several queries to aggregate and slice the data in an appropriate manner. We calculated the sales per product as well as the on-air minutes per product. It’s important to note that we used a left join when joining the on-air minutes; meaning if there is no on-air marketing the observation is still included in the analysis. We believe it is important to include these records and examine the effects of having zero TV time.

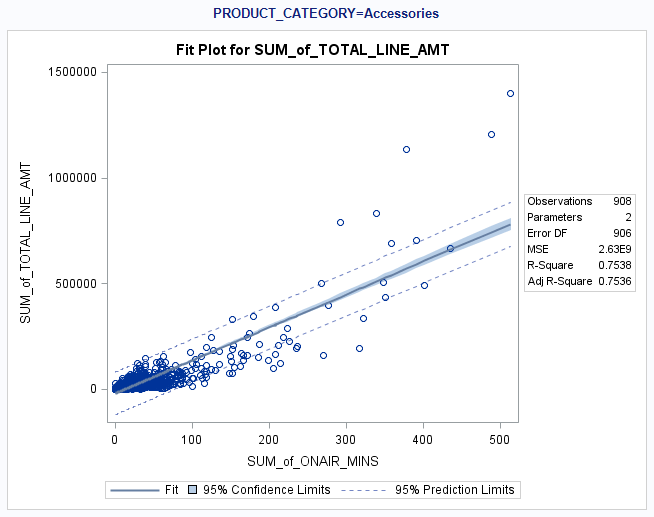
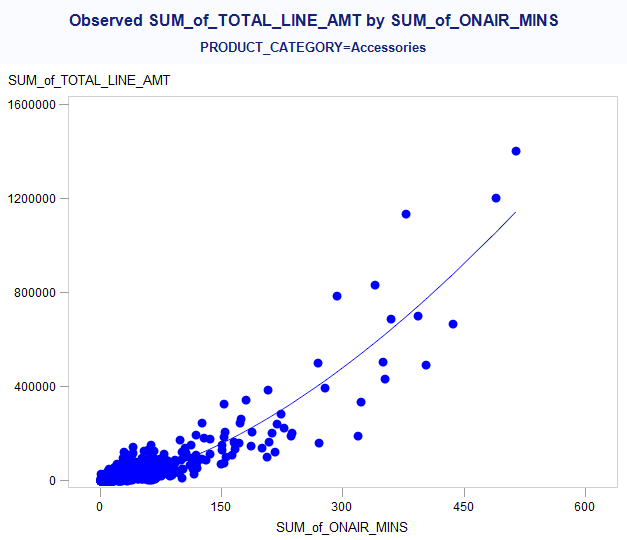
Next, we ran a regression to see whether on-air minutes is a reasonable explanatory variable for the total sales per the product. As expected, there is a very strong positive relationship. With an intercept of -$4,195, each additional 1 minute of air-time can reasonable account for $1,285 of additional sales. To put this into perspective, a product that receives 1 hour of on-air time would project sales to be ~$72K.



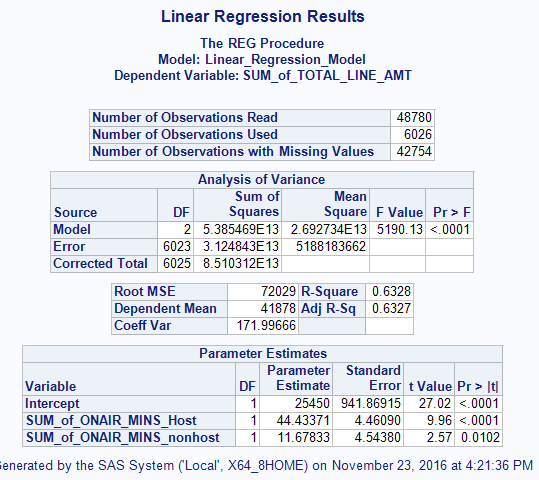
The statistical evidence supports that we can be quite confident in these results, as the standard error of the parameter estimate is only $7.93 (less than 1% of the parameter estimate). Also, the r^2 is .76, meaning approximately 76% of the variation in sales can be reasonably contributed to on-air time.

We also performed an analysis to check the difference in fit/parameters/intercepts for the different product categories. For example, it’s important to know whether on-air time is more efficient at gaining sales for certain product groups. A frequent problem that many companies face these days is prioritization, and specifically determining which product to spend their resources on. In this case, we wanted to partition the regression analyses into the product categories where the product categories had at least 100 observations. Results found that many of the product categories do have quite different intercept and parameter estimates. In particular, Accessories ($1,559), Apparel ($1,507), Health/Beauty (1,412), Home Décor ($1,512) are all very efficient users of on-air time as the per minute increase in sales is relatively high. In contrast, Textile/Furniture ($925) is relatively low and inefficient way to spend the broadcast time.

Visually assessing the Accessories category we recognized a non-linear relationship. After assessing several different non-linear models, we introduced a model with excellent fit to the seemingly exponential curve. Using the Guass-Newton method and the following computation (sales = 48.82 \* OnAirMinutes \*\* 1.61), we discovered a much better fit for the products who fit the category of Accessories (which accounts for a large chunk of sales). This model suggests that for accessories might have an ever-increasing benefit for increased on-air minutes.

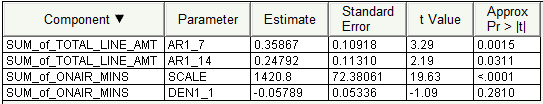


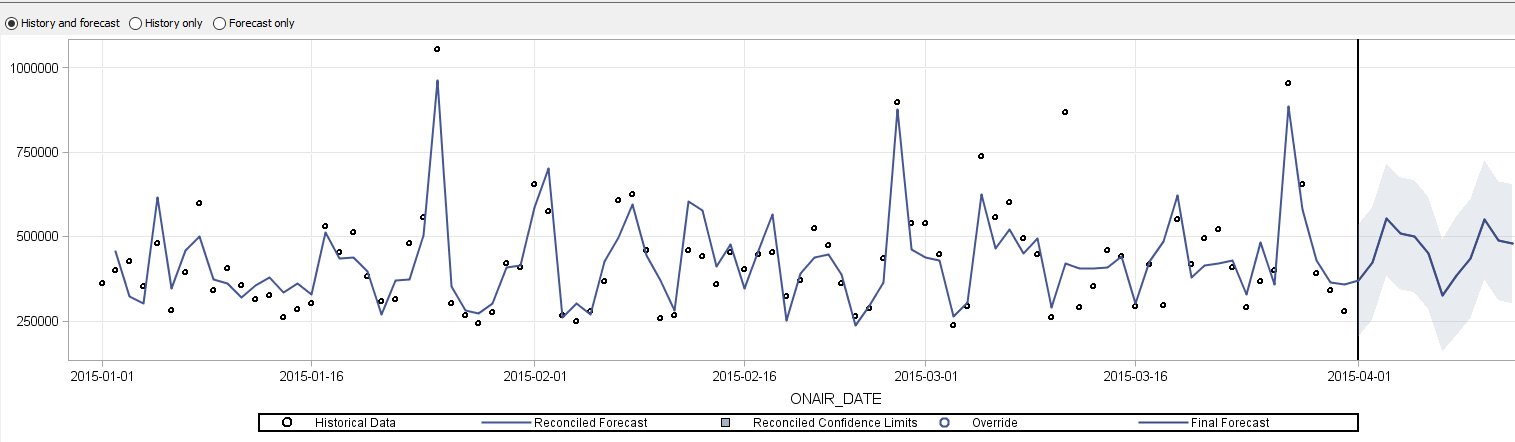
Now that we have established that the product’s time on air is an important driver of sales, we would like to establish whether the hosts make a difference. We found that almost every product that had a host also marketed without a host, so we did not want to use dummy variables here and duplicate our data. We decided instead to perform a similar exercise above (sales = intercept + airtime\*estimate). However, this time we calculated the time the product was on air with a host, and the time the product was on-air without a host (two independent variables). The results are interesting, because the intercept is much higher ($25K) and per-minute parameters are smaller ($44 and $11 for host air-time and non-host air-time, respectively). For interpretation purposes, a product with 200 minutes of host time would have predicted sales of $33,800. On the other hand, 200 minutes of only non-host time would likely generate $27,200. Unless the hosts make more than $33 per minute (~$4 million a year) then it would be wise to have a host as much as possible.



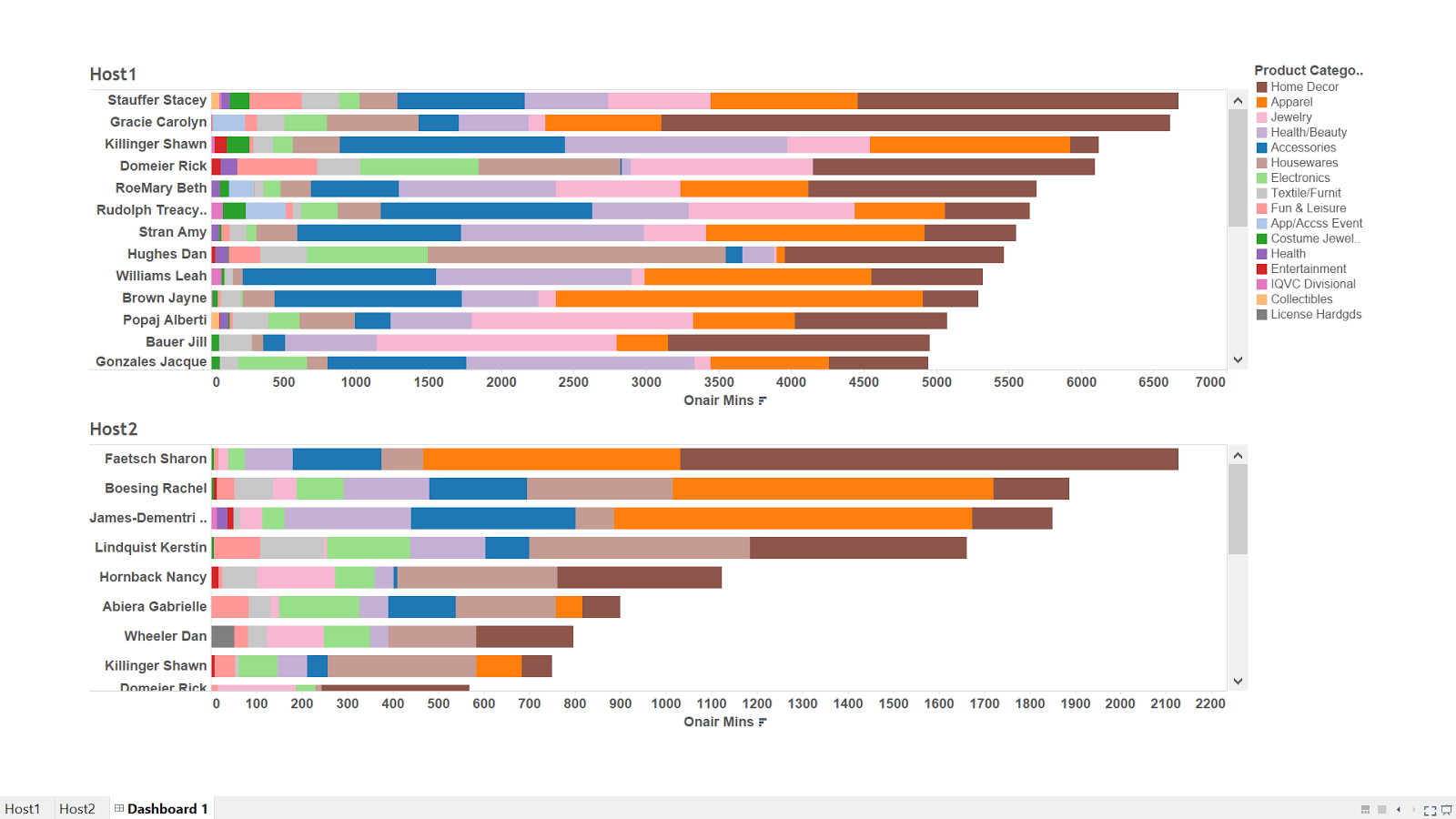
Next, to further validate our assumptions as well as generate a potentially powerful forecasting tool, we generate an ARIMA model in SAS Forecast Studio. Data preparation included generating one dataset with the daily total sales per category as well as the on-air time per category. Forecast studio was grouped by total sales per product category and by day. Product airtime was grouped by product category and by day. It’s important to note due to the fast-paced nature of QVC and hosts’ on-air promotions, we decided to use a daily interval, a weekly seasonal cycle, and on-air time as a potential variable. Also, we categorized the analysis by category.

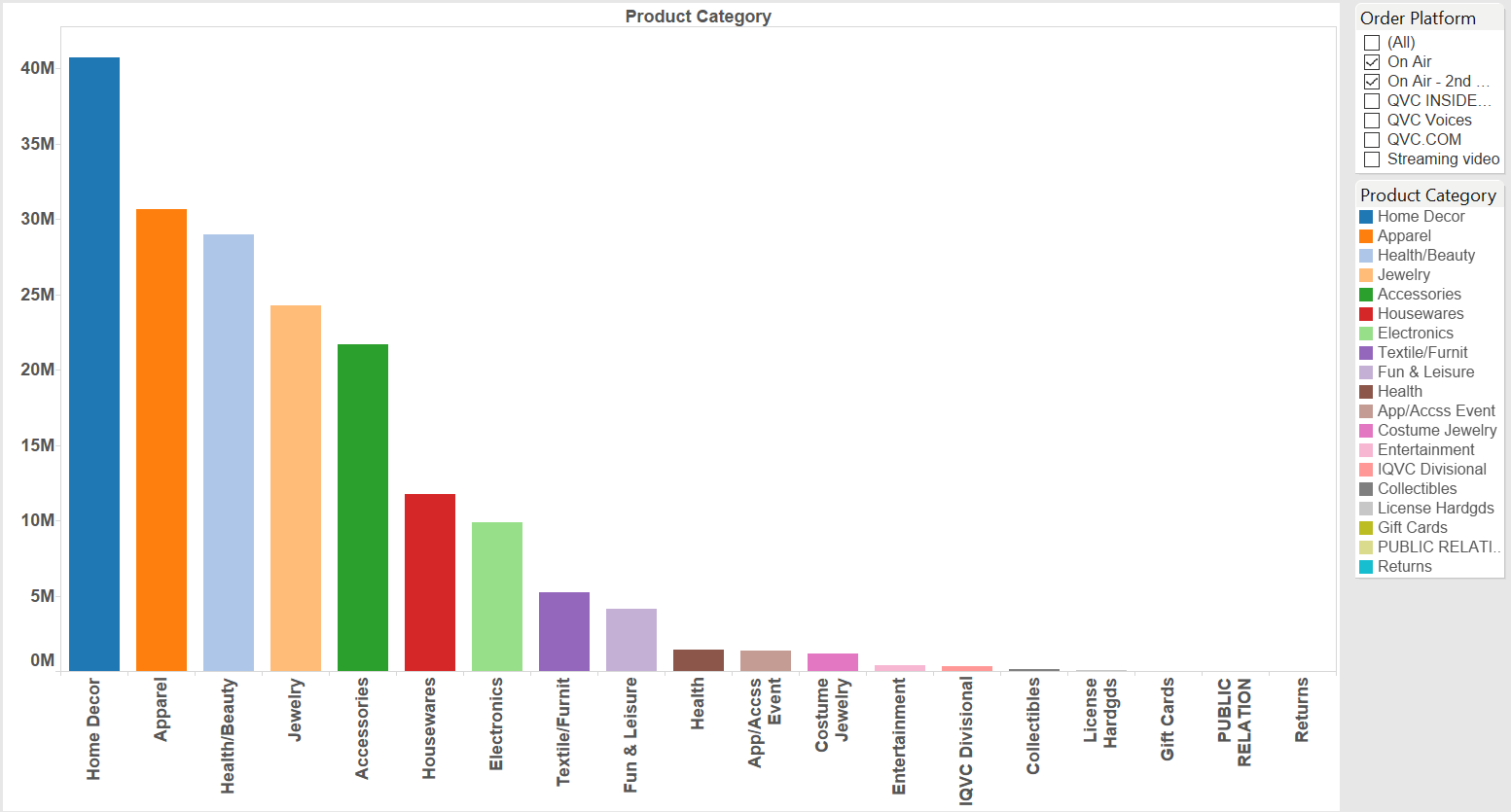
The strongest model using low MAPE is the ARIMA Model with the below pictured parameters. As you can see there are two auto-regressive terms, meaning that the sales from 7 and 14 days ago is important in determining the predicted sales. Also, very importantly, per on-air minute increases the sales estimate by $1,420. This all but validates our assumption in the linear model in terms of a positive relationship that is relatively large.





We also created a visualization in Tableau that depicts the on-air minutes for Host 1 and Host 2. Host 1 will cover more on-air minutes than the Host 2. The average on-air minutes for the Host 1 would be 3000 minutes in a quarter.  But the average on air minutes for the Host 2 would be less than 1000 minutes per quarter.  The highest on air minutes was done by Stauffer Stacey, 7000 minutes.  The highest on-air minutes for a product category is Home Décor. And the highest sold product category is also Home décor because it has the highest on-air minutes.  So on air minutes has direct relation with the products sold.

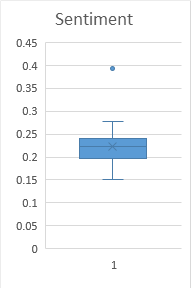
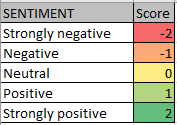




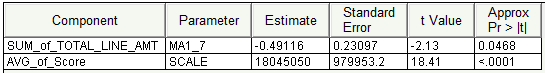
**Product sales revenue is directly proportional to on air minutes.**

**Question 4: What is the impact of QVC’s social networking presence on driving product sales?**

Social network is a very new and interesting topic and can bring significant impacts to business decisions. QVC data for social media comments has been provided for the months of March. Unfortunately, it’s not possible to translate the data to a per product analysis due to the nature of the data sources, however, we can collect the sentiment of each and compare these to sales. To complete this analysis, we looked at sales per day vs average sentiment per day. Since this is a time series approach, we decided to leverage SAS Forecast Studio’s powerful capability in this regard. It’s important to note we translated sentiment of each social media point to an ordinal variable as noted below.

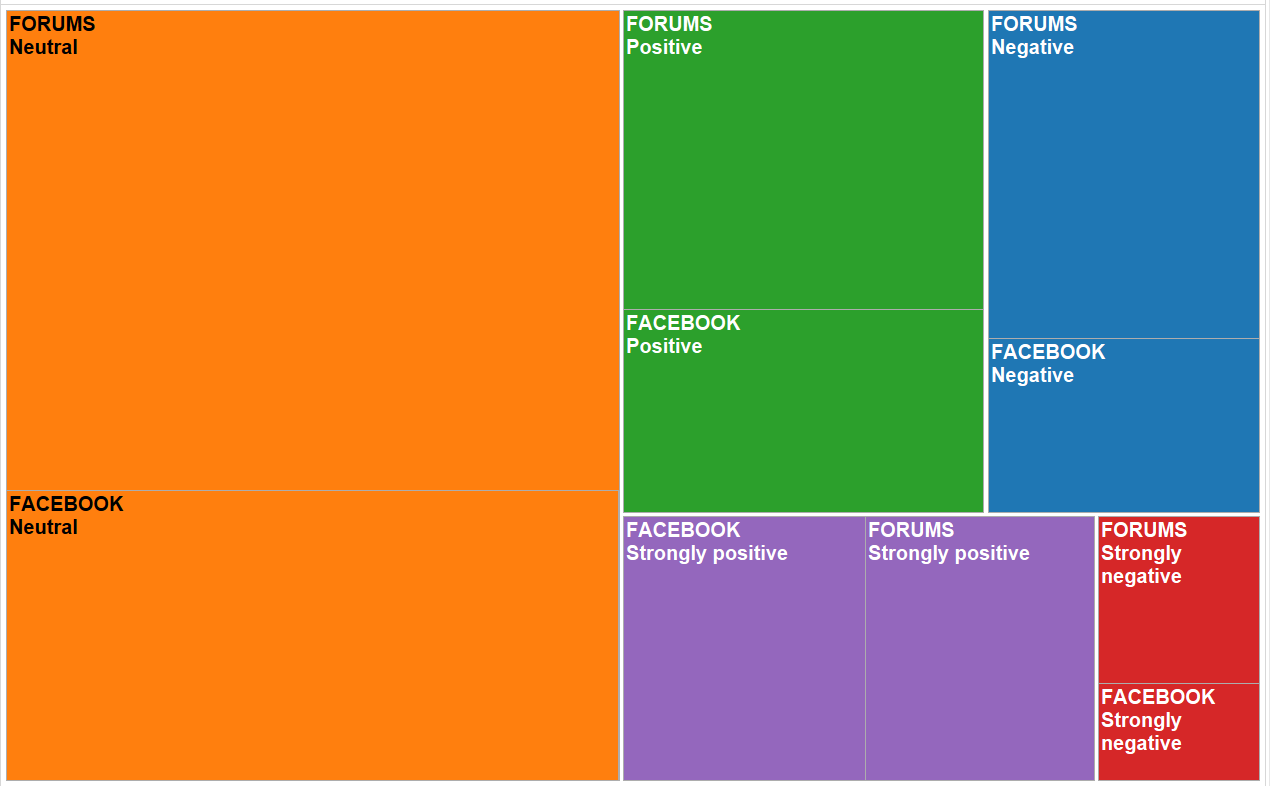


This relatively simple approach turned out to be a significant finding: the more positive the sentiment the more sales (on a daily basis). The ARIMA model chosen has a relatively low MAPE (second lowest) and includes the sentiment score as a significant variable. The model is unlike many ARIMA models, in that it has a negative moving average term (7 days). After multiplying the 7 day sales amount by -.49, $1.8 million is added for each sentiment point increase. To put this into perspective, we can use the sentiment box and whisker to the right. For example, a day with sentiment on quartile one and a day with sentiment on quartile two would have an estimated sales difference of over $200K.



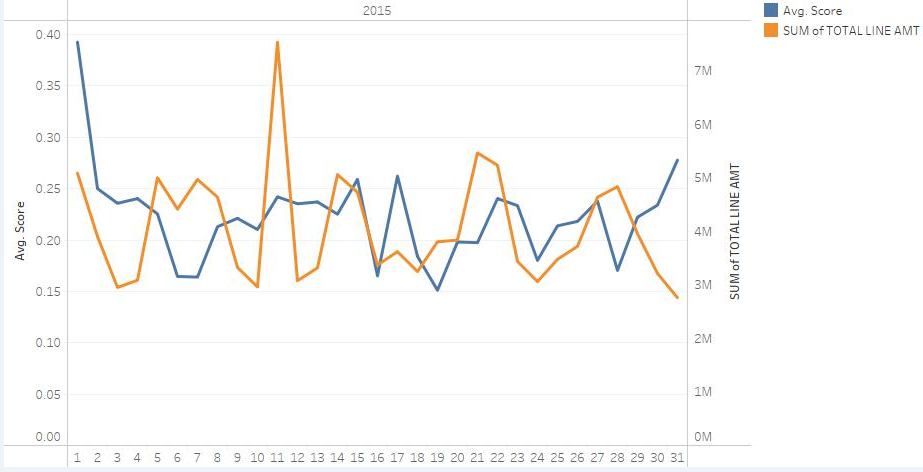
In summary, we suggest being suspect of the above model due to the fact that there is a negative MA term and such a high parameter assigned to the average sentiment. However, it’s clear there is a positive relationship between overall sentiment of the day and total sales for the day. QVC should absolutely try to build more datasets around overall customer satisfaction and move toward a solution for preventable unsatisfied customers.

**Distribution of Sentiments across platforms**



Most of the comments were posted on Facebook and other forums. Approximately 50% of our sentiment is neutral. Overall, less than 20% of the comments had negative or strongly negative sentiment. From this we can derive that overall customers are happy with QVC and their products.

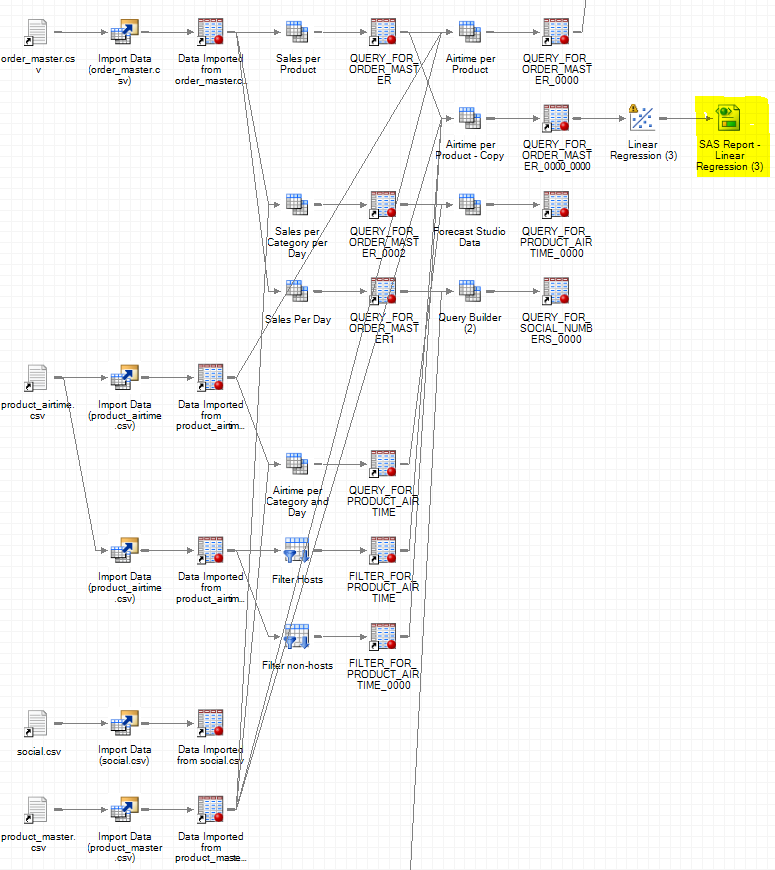
**Sales per day vs Average Sentiment score**



The above diagram depicts that there is a bit correlation between the average sentiment score and products sold during that day.

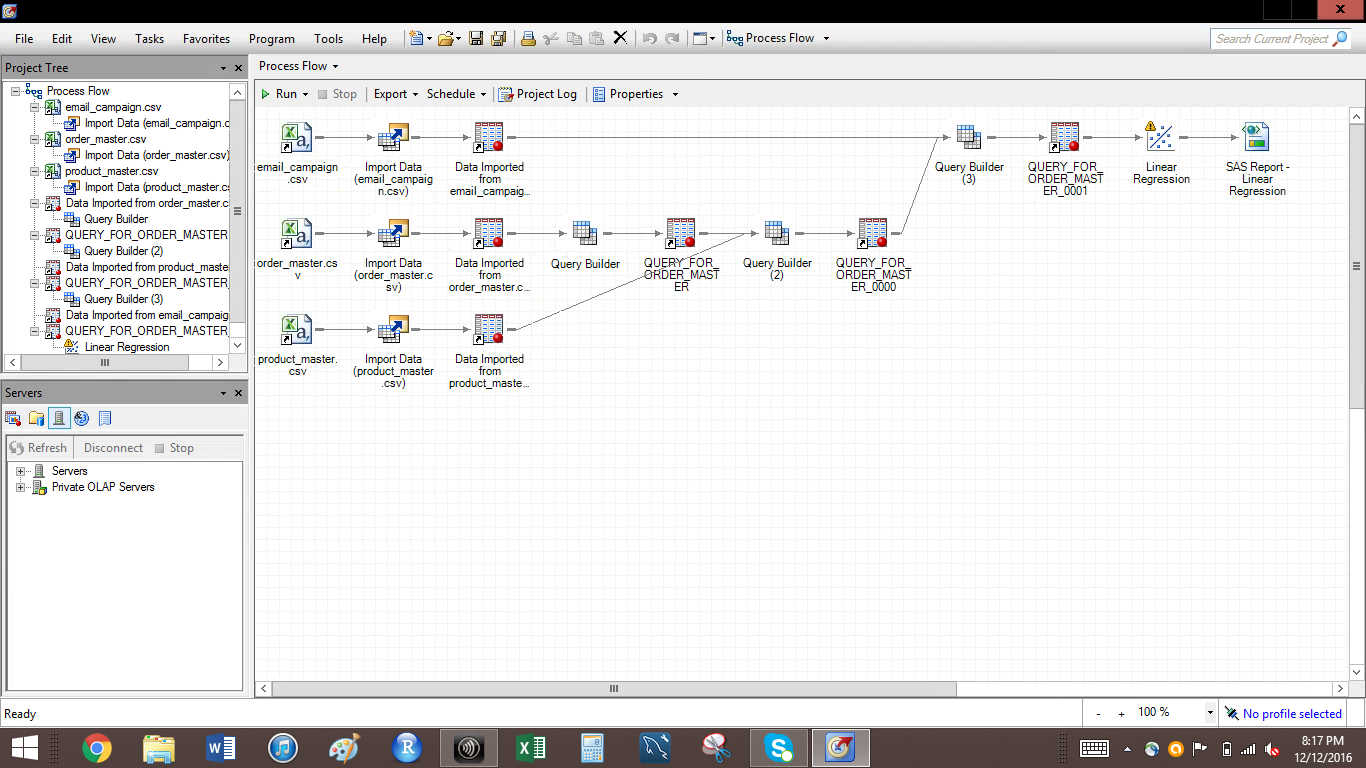
**Process Flow**

**Questions 1, 3 and 4**

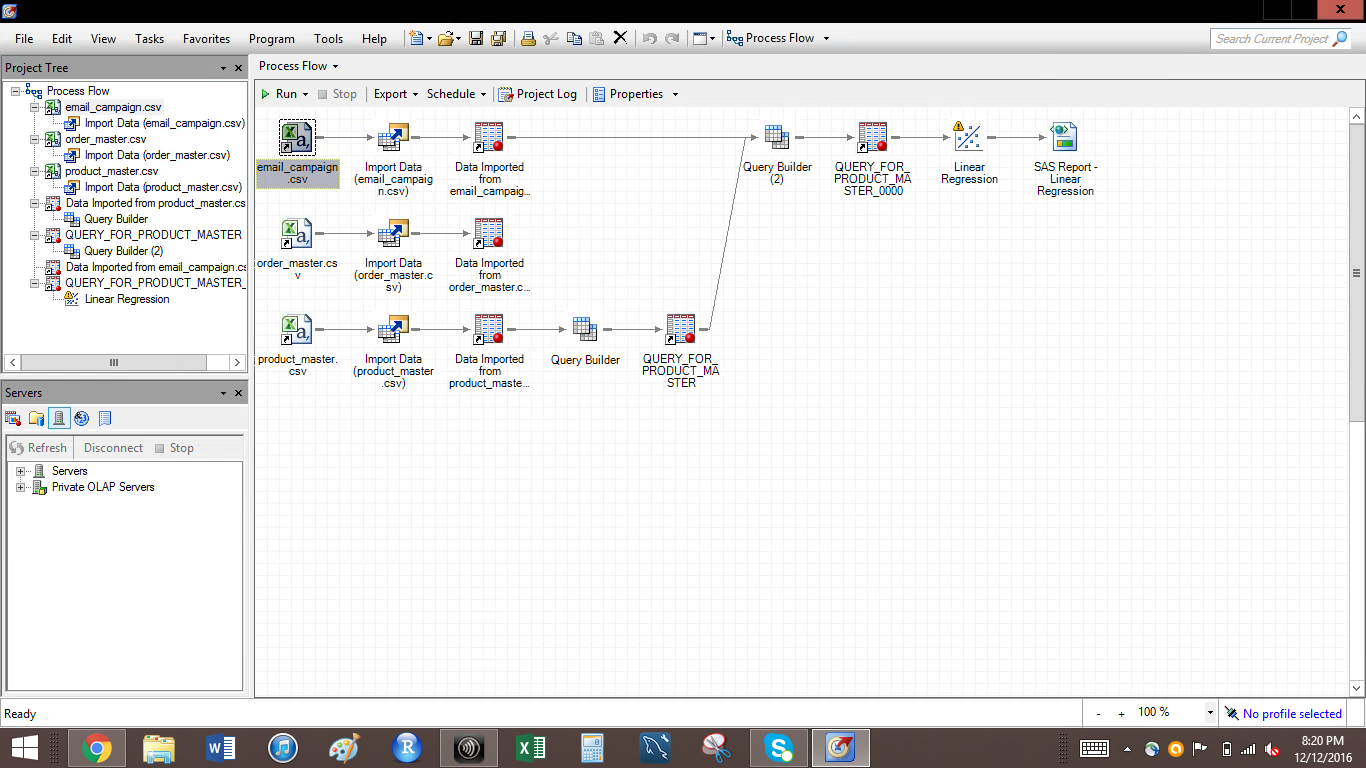


**Process Flow for Question 2**

**Email vs. Total Line Amount flow**



**Email vs. number of products**



**Final Recommendations**

**1) Offers & promotions during peak sales hours**

Since there are more sales during morning and at night, more offers and promotions need to be applied to increase the sales.

**2) More on-air minutes by famous host to increase the product sales**

As there is high correlation between product sales and host on air mins, famous host should be given more on air mins to promote a product to increase its sales.

**3) Focus on the social media comments**

Social media also affects the sales as more positive comments leads to more sales, so the feedbacks which are negative need to be given more importance and improve their quality to increase the sales.