Bryant, Tiana

Wide world importers | CA, USA

sales performance analysis PROJECT PLAN

Table of Contents

[Project Background 2](#_Toc101711576)

[Project Plan 3](#_Toc101711577)

[Week 1 3](#_Toc101711578)

[Week 2 3](#_Toc101711579)

[Week 3 3](#_Toc101711580)

[Week 4 3](#_Toc101711581)

[Week 5 4](#_Toc101711582)

[Week 6 4](#_Toc101711583)

[Week 7 4](#_Toc101711584)

[Week 8 5](#_Toc101711585)

[Project Analysis 6](#_Toc101711586)

[Business Questions 6](#_Toc101711587)

[Preliminary App Development 13](#_Toc101711588)

[Final App Screenshots 16](#_Toc101711589)

[Findings 27](#_Toc101711590)

[Proposal 1 28](#_Toc101711591)

[Proposal 2 28](#_Toc101711592)

[Proposal 3 28](#_Toc101711593)

[Conclusion 28](#_Toc101711594)

**Project Background**

**Analytic Theme:** Sales Performance

**Required Analyses:** Sales by Salesperson, Territory, Profits, Calculated Costs

**Business Process:** Sales

**Logic:** To explore sales reporting that includes territory, salesperson, customer, and supplier. This can aid the business to determine who frequent customers, frequent suppliers are and the products that are typically purchased. Additionally, the analysis will cover the profits associated with the sales that are made. The analysis can also help identify most profitable sales territories to determine where to make strategic efforts to market in those areas.

The business desires to have basic sales reporting that shows the orders by current territory, by rank, so that the sales reps can place their focus on the customers ranked by orders. This query extracts the data from the sales, customer, and territory tables to produce a report that shows the customer’s total orders, ranked by those orders, to help sales reps identify the top customers in each region and group.

**Location of Files:** <https://github.com/bryantt11/wwi.analysis>

## **Project Plan**

This project plan provides an overview of the activities over the course of the project.

|  |
| --- |
| **LEGEND** |
| Not Started |
| In Process |
| Completed |

### Week 1

* Identify data sources and location
* Confirm access to data

### Week 2

* Data exploration
* Identify business questions
* Extract data
* Load data to R
* Provide project status 1 via check-in

### Week 3

* Transform data to prepare for analysis
* Perform analysis of business questions
* Develop draft dashboard ideas (R Shiny)

### Week 4

* Final project plan
* Begin with dashboard
* Provide project status 2 via check-in

### Week 5

* Determine visualizations to be used
* Develop model for R
* K Means or Louvain?
  + To identify clusters within the dataset, two methodologies were considered: K Means and Louvain. K means and Louvain are both adept with unstructured data to identify associations within the dataset. The Louvain algorithm is similar and, using Neo4j and Cypher, can identify clusters within a network. The purpose of identifying clusters would be to identify outliers within the dataset as well as to identify problems that hadn’t considered.
  + Given the size of the dataset and a lack of skill development in Neo4j, this would have been a quite complex task in either scenario. The decision was made to utilize K Means in R due to the familiarity of the tools available.

### Week 6

* Draft user instructions
* Finalize visualizations
* Provide project status 3 via check-in

### Week 7

* Document findings
* Dashboard testing
* Finalize user instructions
* Provide final project status via check-in

### Week 8

* Create presentation
* Present findings
* Deploy app
  + I could not figure out the issue that prevented the app from deploying. The app is currently under test and once complete will be deployed using shinyapps.io and posted here: <https://bryantt11.github.io/TianaBryant.github.io/>

## **Project Analysis**

### Business Questions

#### Descriptive

##### Are sales increasing or decreasing by year?

Graphical user interface, table

Description automatically generated with medium confidence Chart, bar chart

Description automatically generated

##### What are the top 5 sales territories?

Table

Description automatically generated Chart, bar chart

Description automatically generated

##### Who are the top 5 salespeople for WWI?

Text

Description automatically generated with low confidence Chart, bar chart

Description automatically generated

##### Which sales territories are the most profitable?

A picture containing table

Description automatically generated

##### Who are the top suppliers?

Graphical user interface, text

Description automatically generated

##### What are the top products sold?

A picture containing text, receipt

Description automatically generated

##### What kind of products are commonly sold?

Graphical user interface, text, application

Description automatically generated

##### Where are the 3 least profitable, and what is the costs associated with doing business in these territories?

Table

Description automatically generated

##### Which of WWI’s suppliers are associated with the most profits? Why are the ones that do not have profitability performing poorly?

Table

Description automatically generated

##### What is ROI on products purchased by WWI to sale?

Text

Description automatically generated

* Correlation plot to identify key correlations and multicollinearity.

Chart, scatter chart

Description automatically generated

#### Predictive

##### Linear model

To understand the common occurrences contained within the dataset, a linear regression model was created. The response variable was Profit and the remaining variables the predictor. The results a few variables with great statistical significance, including Month, Calculated Unit Cost and Stock Item.

Table

Description automatically generatedChart, line chart

Description automatically generated

##### Stepwise regression

To get a better idea of what variables might be the most important in the dataset, stepwise regression was used. This was done to identify which independent variables might most influence the dataset.

Table

Description automatically generatedTable

Description automatically generated with medium confidence

A picture containing table

Description automatically generated

Text

Description automatically generated

\* ME, RSME, MAE, MPE and MAPE

A picture containing chart

Description automatically generated A picture containing chart

Description automatically generated A picture containing chart

Description automatically generated

##### K Means Clustering

Though this data was highly structured, there was a need to explore further to see if there were any other associations within the dataset that had not yet been caught. Using K Means helps to find those associations.

* Originally selected 2, decided on 6 using the elbow method. In looking at the results, it looks like cluster 4 appears to be the outliers/catch all cluster or represent the sales that were low quantity, etc. Higher profits and costs are associated with cluster 3, higher quantities in cluster 5.

Chart, line chart

Description automatically generated Chart

Description automatically generated

Table

Description automatically generated

A picture containing graphical user interface

Description automatically generated

##### Neural Networks

Used predictors identified in stepwise regression (both) and considered multicollinearity.

* In using a neural network, patterns can be detected and help identify issues and improve making within the business. Using some the predictors found from the stepwise regression exercise.

Text

Description automatically generated Diagram

Description automatically generated

Text

Description automatically generated Diagram

Description automatically generated

## Preliminary App Development

This is an overview of the application during the development phases. Screenshots are included of the progress thus far.

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Graphical user interface, application

Description automatically generated

Graphical user interface

Description automatically generated with low confidence

Graphical user interface

Description automatically generated

## Final App Screenshots

After further reviewing the app, it was decided to add some selections to the sample criteria so that the user could select the region as well. Also added was a KNN model to the list of predictive models used in the project. The below show the app in its final state.

Application

Description automatically generated with low confidence

Graphical user interface, text, application

Description automatically generated

Chart, bar chart

Description automatically generated

Chart, bar chart

Description automatically generated

Graphical user interface, application

Description automatically generated

Graphical user interface, application, table

Description automatically generated

Graphical user interface

Description automatically generated with low confidence

Graphical user interface

Description automatically generated with medium confidence

A picture containing graphical user interface

Description automatically generatedGraphical user interface

Description automatically generated

Graphical user interface

Description automatically generated

Graphical user interface

Description automatically generated

A picture containing graphical user interface

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application

Description automatically generated

## Findings

Based on the historical data previously analyzed, what can be said about future profitability? Costs? Using various methods of exploring the dataset through the predictive models developed, it was discovered that there were a few key predictors of profit. It can be determined unit costs have a significant relationship to profits. Consistently throughout the analysis, the as unit costs went up, so did profits. This influenced the analysis and the focus narrowed to how costs impact sales and, ultimately, profits.

While we have limited data from 2016, it appears that sales are consistently growing year-over-year. The most profitable regions over the four-year lookback are Southeast ($34MM), Mideast ($23MM) and Southwest ($22MM). Top performers for all regions include Amy Trefl, Anthony Grosse, and Archer Lamble.

Briefly, it may appear that the External region performs superiorly with a mean profit of $383. When diving deeper, the costs associated with doing business are the highest of all the regions with a mean at $24, exceeding the other regions by nearly a dollar or more. Other low profit regions include New England and Rocky Mountain.

In terms of suppliers, it appears that Litware, Fabrikam and Northwind Electric Cars are the largest suppliers contained regarding sales within the dataset. The top products sold include Stock Item Keys 5 (Air Cushion Machine), 145 (Ride On Big Wheel Monster Truck) and 146 (Ride On Vintage American Coupe) respectively. All these products fall within the Novelty Shop product type.

### Proposal 1

Reconsider doing business in the External, New England and Rocky Mountain regions, given the cost relative to profit exceeds top performers.

### Proposal 2

Consider shifting sales strategy to utilize suppliers whose products cost less compared to profit. For example, while Northwind Electric Cars had 2 out of the top 3 in terms of products sold, the cost relative to profit is 26.1%. This means that for every $1 profit, there is more than 26 cents to earn that profit. Litware, which held the number 1 sales spot, is 5 cents for every dollar. The KNN model also supports this proposal, as this supplier's involvement in each transaction is a likely predictor of a sale.

### Proposal 3

Consider eliminating doing business with A Datum Corporation and Contoso. While costs are relatively low, there is not much profit gained in selling products from these two suppliers. Consider replacing or shifting efforts to suppliers whose products produce a higher yield.

### Conclusion

Currently, ROI stands at 884%, which is quite respectable. If cost cutting measures associated with the proposals take place, the business might find that the efficacy of the ROI improves and ROI increases.