**Course Project Overview**

**Directions:** Review the information below to learn about the specific content required for each component of the course project and the timeline for completing each component.

**Background Scenario:**

You are a senior data analyst. You work for an independent analytics firm, contracted by AdventureWorks, to provide insights into their operations. In CPC Assignment #1, you will provide some basic discovery analytics about their sales. In Assignment #2 you will enhance your presentation to include a variety of visualizations. Finally in Assignment #3, you will provide some predictive analytics. Each assignment will require you to use R connected to SQL Server. To learn how to use data in R that is sourced from a SQLServer database, refer to the course materials. Information is also widely available on the Internet. As soon as possible, read through all three assignments so that you may incorporate what you learn in lecture and readings into your CPC work product.

**Retrieving the Data:**

Data will come from the AdventureWorks2016 SQLServer database. For CPC Assignments #1 & 2, detail data should be retrieved, aggregations should be performed in R. For CPC Assignment #3, data can be aggregated in SQL or in R.

**Communicating to Stakeholders:**

You will be creating a living document through the three components. As you get your graded work returned, make sure to modify your previous work to avoid duplicate deductions. See the "Guide to the Slides" section below for details on how to assemble your PowerPoint presentation.

**Project Components:**

| **Component Number and Title** | **Topic Content Used to Create Component** | **Component Specifications** | **Component Deliverables / Tasks** | **Due Date** |
| --- | --- | --- | --- | --- |
| **Component 1:**  Initial Data Analysis and Summary Statistics | * How to load data from SQLServer into R structures. * How to generate summary statistics from detail data. * How to manipulate data objects. * How to select and display parts of data objects. * How to sort and rearrange data objects. * How to summarize data samples. | * Identify Top 10 products by both sales revenue and unit volume for orders with an order date in 2013 * Present these two lists in tabular format with appropriate information. * For all orders in 2013, identify relevant summary statistics (max, min, mean, median, etc.). * Answer the Component Questions at the end of this document | Complete slides 1-5:  **Item 1**: Title slide  **Item 2:** One overview slide  **Item 3:** One slide with tables including data for the Top 10 Products by Unit Volume  **Item 4:** One slide with tables including data for the Top 10 Products by Sales Revenue  **Item 5:** Summary statistics for **all** 2013 sales orders.   * In the Word document, include the SQL and R code used to complete this assignment. * In the Word document answer the Component Questions. | End of Topic 3 |
| **Component 2:**  Cumulative Statistics and Data Analysis | * How to convert tabular data to visual representations * How to communicate a message with data visualizations | * Convert tabular data for each Top 10 list (sales and volume) into a pie chart and a bar chart. * Create line charts with 2013 sales data by month. * Create elements that make visualizations more understandable, including use of legends, colors, and text labels. * Answer the Component Questions at the end of this document | * Complete slides 6-7: **Item 1:** Visual representation (pie and bar charts) of Top 10 products by Unit Volume.   **Item 2:** Visual representation (pie and bar charts) of Top 10 products by Sales Revenue.   * Complete slides 8-10:   **Item 3:** A line chart visualizing sales by month for the number one product by Sales Revenue.  **Item 4:** A line chart visualizing sales by month for the number two product by Sales Revenue.  **Item 5:** A line chart visualizing sales by month for the number three product by Sales Revenue.   * In the Word document, include the SQL and R code used to complete this assignment. * In the Word document answer the Component Questions. | End of Topic 5 |
| **Component 3:**  Graphical Analysis, Linear Regression, and Recommendations | * How to create a range of graphs to summarize data and results. * How to create scatter plots, including a multiple correlation plot. * How to perform linear regression (including multiple regression). * How to plot regression models. * How to add lines of best-fit to a regression plot. | * Using 2013 data for the top 3 products by Sales Revenue, generate predictions for Sales Revenue for the first six months of 2014. Order month is the independent variable. * Plot these predictions on a line chart and identify the predictions using a different point type and color. Include the line-of-best-fit (abline) * Create elements that make visualizations more understandable, including use of legends, colors, and text labels. * Explain how to interpret the linear model equation * Answer the Component Questions at the end of this document | * Complete slides 11-13:   **Item 1:** A line chart visualizing sales by month for the number one product by Sales Revenue.  **Item 2:** A line chart visualizing sales by month for the number two product by Sales Revenue.  **Item 3:** A line chart visualizing sales by month for the number one product by Sales Revenue.   * Include linear regression predictions for the first 6 months of 2014 for items 1-3. * **Item 4:** Provide recommendations to the client regarding products based on the predictions you made. * In the Word document, include the SQL and R code used to complete this assignment. * In the Word document answer the Component Questions. | End of Topic 7 |

2013 Top 3 Products by Revenue:

|  |  |
| --- | --- |
| **Product Name** | **Revenue** |
| Mountain-200 Black, 38 | $2,212,974.78 |
| Mountain-200 Black, 42 | $1,932,388.29 |
| Mountain-200 Silver, 38 | $1,815,673.09 |

2013 Top 3 Products by Unit Volume:

|  |  |
| --- | --- |
| **Product Name** | **Unit Volume** |
| Water Bottle - 30 oz. | 3,913 |
| AWC Logo Cap | 3,768 |
| Sport-100 Helmet, Black | 3,088 |

## Guide to Slides

Use the table below to guide the development of each slide. You are required to include relevant speaker notes for each slide. See “Guide to Slides” section in this document

|  |  |  |
| --- | --- | --- |
| **Slide Number** | **Name** | **Description** |
| **Week 3: CPC Component 1** | | |
| 1 | **Item 1:** Title Slide | Your name, title, company, client name, date, and presentation title |
| 2 | **Item 2:** Overview | Briefly restate the purpose of the project, include pertinent information about the data (source, date range, quality issues if any) |
| 3 | **Item 3:** Top 10 Products by Unit Volume | Includes table containing information about the top 10 products by Unit Volume (SalesOrderDetail.QtyOrdered) |
| 4 | **Item 4:** Top 10 Products by Sales Revenue | Includes table containing information about the top 10 products by Sales Revenue (SalesOrderDetail.LineTotal) |
| 5 | **Item 5**: Summary Statistics | Summary Statistics for the order amount for **all** orders for 2013 (SalesOrderHeader.TotalDue) |
| **Week 5: CPC Component 2** | | |
| 6 | **Item 1:** Charts for Top 10 Products by Unit Volume | A pie chart and a bar chart visualizing the Top 10 products by Unit Volume (visual version of Slide 4) |
| 7 | **Item 2:** Chart for Top 10 Products by Sales Revenue | A pie chart and a bar chart visualizing the Top 10 products by Sales Revenue (visual version of Slide 3) |
| 8 | **Item 3:** 2013 Sales by order month for #1 Product by Sales Revenue | A line chart visualizing 2013 monthly sales for the #1 product by sales revenue. x-axis should be order month, y-axis should be monthly sales revenue. |
| 9 | **Item 4:** 2013 Sales by order month for #2 Product by Sales Revenue | A line chart visualizing 2013 monthly sales for the #2 product by sales revenue. x-axis should be order month, y-axis should be monthly sales revenue. |
| 10 | **Item 5:** 2013 Sales by order month for #3 Product by Sales Revenue | A line chart visualizing 2013 monthly sales for the #3 product by sales revenue. x-axis should be order month, y-axis should be monthly sales revenue. |
| **Week 7: CPC Component 3** | | |
| 11 | **Item 1:** 2014 Sales Predictions for #1 Product by Sales Revenue | A line chart visualizing 2013 monthly sales for the #1 product by monthly sales revenue including predictions for January-June of 2014. |
| 12 | **Item 2:** 2014 Sales Predictions for #2 Product by Sales Revenue | A line chart visualizing 2013 monthly sales for the #2 product by monthly sales revenue including predictions for January-June of 2014. |
| 13 | **Item 3:** 2014 Sales Predictions for #3 Product by Sales Revenue | A line chart visualizing 2013 monthly sales for the #3 product by monthly sales revenue including predictions for January-June of 2014. |
| 14 | **Item 4:** Recommendations | A slide containing at least three recommendations about products based on the predictions for future sales revenue. |
| END | End Slide | A simple closing slide to display when taking questions |
| REF | References | Any References in APA style (Optional) |

## Component Questions (Complete on a Word document)

CPC #1

1. Any initial thoughts based on your initial discovery?
2. What are the advantages/disadvantages to presenting tabular data as opposed to visual data?

CPC #2

1. Thinking about sales of exercise equipment. Would sort of trends might you expect to see? For instance, any periods of increased/decreased sales?
2. Comment on the month-to-month variability of sales for each of the three products.
3. Do you visually see any trends for the three product monthly sales?

CPC #3

1. Why might linear regression be a bad idea for forecasting sales when we only have one-year worth of data?
2. Do you believe these models could be used for any future date or is there likely a limit to how far out you can reliably predict?
3. List the linear model equations for each of the three products. Provide an example of how to generate a prediction using these equations by hand.

## Submission

You will need to submit two files:

1) A PowerPoint file (pptx) containing your presentation.

2) A Word document (docx) with your R and SQL code and answers to the discussion questions.

Note that R and SQL code should not appear in the PowerPoint presentation.

## Sample Code for Linear Regression Activity in CPC Assignment #3

Text

Description automatically generated