**Team 3**

**Data Analytics Interim Project Proposal**

**Overview**

As a team, we have discussed each question to first decide how we would approach it, any clarifying questions we may have, and then finally discussing the best visualisation tools that would be used to reflect and portray any insights. Once this was completed, each member had the opportunity to share their points of view, and so all perspectives regarding how to tackle the question were considered. We then divided each question between the members, taking into account how vocal they were in giving an opinion on how to answer the question, as this showed they had a good enthusiasm and positive outlook in taking it on.

We then start working on our questions independently, but can still ask questions to the team in order for us to be able to progress. After finishing our work, we will go through each person’s work and give constructive advice/criticism.

Finally adding our slides into a collaborative Powerpoint presentation using Google Slides, we will order the responses to each question to convey a story of our insights and recommendations.

|  |  |
| --- | --- |
| Proposed by: | Team 3  Ruzzel Pescozo - Team Leader  Charles Clarke  Basil Wahid  Sava Drobot  Silk Somalin |
| Timeframe: | Completion by 11/09/2023  Presentation on 11/09/2023 |
| Instructor: | Griffin John Maree |

**Objectives and steps**

With the use of specific data found in Adventure works we were tasked to answer the following questions:

1. What is the regional sales in the best performing country?

Assigned to: Charles

Look at relevant tables (in the Sales schema)

* SQL query for highest sales and check the region
* Pie Chart, visualisation, Comparing country sales
* Part 2, Bar Graph, show regional sales for the best performing country (USA)
* Part 3, why are they the best performing country? (types of products, sales reason)

Look at best performing country in total (not per capita)

1. What is the relationship between annual leave taken and bonus?

Assigned to: Basil

Look for annual leave and bonus column, “vacation hours” in HumanResources.Employee -> link through keys to “bonus” in Sales.SalesPerson

* SQL query to only get annual leave and bonus, we don’t care about who took the leave and who got the bonus
* Create Scatter Plot visualisation for the different questions i have after i have created a scatter for vacation hours and bonus
* Created multiple scatter graphs which relate sales to bonus and vacation hours, as well as sales to commission

1. What is the relationship between Country and Revenue?

Assigned to: Basil & Charles  
Charles - SQL, Basil - SQL & Python

* Use SQL to create table group by countries, columns for sales revenue and total number of sales (Charles)
* Use SQL to create table which counted all the stores by the country code and then created views using both of the queries (Basil)
* Use python to Create a horizontal bar graph which has the countries and their revenues andUse python to create scatter of Number of stores vs Revenue, with each point being a different country (Basil)

1. What is the relationship between sick leave and Job Title?

Assigned to: Sava

SQL statements for Sick leave hours, Job Title and Last Department, organisation level

Histogram of Sick Leave hours, descriptive statistics

Box plots of Sick Leave Hours by departments.

Box Plots of Sick Leave Hours by organisation level

Box plots of job titles of first and 4th quartile. .

1. What is the relationship between store trading duration and revenue?

Assigned to: Silk

-To Solve i used SQL query to only get Trading Duration from Year opened and Annual revenue from sales.vstore with demographic view table

-And then created bar plot to show the relationship, to be more brief i used Scatter Plot to visualise and look for insight such as correlation between two factors

-Can also add more visualisations if needed

1. What is the relationship between the size of the stores, number of employees and revenue?

Assigned to: Ruzzel  
  
Clarifying Questions asked:  
1.What do you mean by “size of the stores”? A: Literal size (sqr ft)  
2.Number and Revenue values are for each store? A: Yes

Potential visualisation = scatter plot with varying circle size (revenue)

1.Scatterplot(x = store size, y = number of employee, S = revenue)

or

2.Scatterplot(x = employee number, y = revenue, s = size of store)

or

3.Scatterplot(x = store size, y = revenue, s = number of employe)  
  
Decided on option 1 for the main visualisation  
Can further experiment with different visualisations to potentially better present insight