**Group 5**

**Data Analytics Interim Project Proposal**

**Overview**

Adventure Works Cycles is a manufacturer and distributor of bicycles and components in the North American, European and Asian commercial markets, and has over 500 employees.

The objective of this project is to analyse the company’s database in order to explore and understand the relationships between various factors, with a focus on store sales and employee performance. By examining these relationships, we aim to gain insights that can inform decision-making and potentially optimise business operations.

We propose to use a combination of statistical analysis techniques and data visualisation methods. Correlation analysis will help identify the strength and direction of the relationships between variables. Additionally, descriptive statistics will provide insights into the distributions of the data.

| Proposed by: | Group 5  Aqsa Shuja  Iris Engler  Stephy Zhu |
| --- | --- |
| Timeframe: | Completion by 25/05/2023  Presentation on 24/05/2023 |

**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?

Data from **Sales.SalesTerritory table**, download the data as csv file and process in Jupyter, use **pandas** to load data and calculate **SalesYTD** **groupby** **CountryRegionCode** first, find the best performing country, then find the regional sales for that country.

Packages: pandas, matplotlib.pyplot, seaborn

Chart: pie chart / bar chart

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

Data from **HumanResources.Employee (column VactionHours)**  and **Sales.SalesPerson (column Bonus), join on BusinessEntityID** to find the bonus and remaining annual leave hours for each sales person.

Packages: pandas, matplotlib.pyplot, seaborn

Chart: Scatter plot / heatmap to see the correlation and distribution

1. What is the relationship between Country and Revenue?

Data from **Sales.vStoreWithDemographics (All columns)** and **Sales.vStoreWithAddresses (CountryRegionName), join on BusinessEntityID.** Find out the total **AnnualRevenue grouped by CountryRegionName.**

Packages: pandas, matplotlib.pyplot, seaborn

Chart: Bar chart and Pie chart

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

Data from **HumanResources.Employee (column JobTitle and SickLeaveHours), HumanResources.EmployeePayHistory(column Rate), join on BusinessEntityID**

Packages: pandas, matplotlib.pyplot, seaborn

Chart: Scatter Plot to see what would happen if we group by Job Title and calculate (average) the sick leave, if doesn’t make sense, we can involve the pay rate to see if there’s a relationship. Use box plot or scatter plot

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

Data from Q3, add a column to calculate years opened, called **OpenDuration**

Packages: pandas, matplotlib.pyplot, seaborn

Chart: **Box plot** to display the distribution of the annual revenue of stores based on their years since opening.

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

Data from Q3, use column **SquareFeet**, **NumberEmployees and AnnualRevenue**. Run correlation test and create a heatmap to showcase.

Packages: pandas, matplotlib.pyplot, seaborn

Chart: Heatmap of correlation, scatter plot to see the distribution.