A logo for college computing

Description automatically generated

**Assessment Cover Page**

|  |  |
| --- | --- |
| *Student Full Name: Cristhian Elson Pereira Macedo* |  |
| *Student Number: 2024104* |  |
| *Module Title: Statistical Techniques for Data Analysis* |  |
| *Assessment Title: CA1* |  |
| *Assessment Due Date:* |  |
| *Date of Submission:* |  |

**Declaration**

By submitting this assessment, I confirm that I have read the CCT policy on academic misconduct and understand the implications of submitting work that is not my own or does not appropriately reference material taken from a third party or other source.

I declare it to be my own work and that all material from third parties has been appropriately referenced.

I further confirm that this work has not previously been submitted for assessment by myself or someone else in CCT College Dublin or any other higher education institution.



**Tittle**

**Cristhian Elson Pereira Macedo**

**2024104**

Higher Diploma in Science in Data Analytics for Business

Statistical Techniques for Data Analysis (StatsDA)

Lecturer: Kayoum Khbuli

CCT College Dublin

Dublin, Ireland

2024



**SUMMARY**

[Table of Figures v](#_Toc164197119)

[1. Introduction 6](#_Toc164197120)

[2. Data Understanding 6](#_Toc164197121)

[2.1 Importing Libraries 6](#_Toc164197122)

[2.2 Loading Data 6](#_Toc164197123)

[2.3 Inspecting the Data Set 6](#_Toc164197124)

[2.4 Checking types of data 6](#_Toc164197125)

[3. DATA PREPARATION 6](#_Toc164197126)

[3.1 Finding Unique Values 6](#_Toc164197127)

[3.2 Replacing Values (Syntax errors) 6](#_Toc164197128)

[3.3 Dropping Irrelevant Columns 6](#_Toc164197129)

[3.4 Renaming Columns 6](#_Toc164197130)

[3.5 Handling Missing Values 6](#_Toc164197131)

[3.6 Imputing Missing Values 6](#_Toc164197132)

[3.7 Handling Duplicate Rows 6](#_Toc164197133)

[3.8 Type Conversion 6](#_Toc164197134)

[3.9 Relationships 6](#_Toc164197135)

[3.10 Statistical Analysis 6](#_Toc164197136)

[3.11 Dealing Outliers 6](#_Toc164197137)

[3.12 Creating Dummies 6](#_Toc164197138)

[3.13 Scaling / Transformation 6](#_Toc164197139)

[4. MACHINE LEARNING - UNSUPERVISED LEARNING 7](#_Toc164197140)

[4.1 Principal Component Analysis (PCA) 7](#_Toc164197141)

[4.2 k-Means Clustering 7](#_Toc164197142)

[4.3 Creating examples to predict 7](#_Toc164197143)

[4.3.1 Preparing for analyses 7](#_Toc164197144)

[4.3.2 k-Means Clustering 7](#_Toc164197145)

[5. DATA DICTIONARY 7](#_Toc164197146)

[References 7](#_Toc164197147)

# **Table of Figures**

# **Introduction**

A objective, problem definition, scope, data sources, ethical considerations, etc

# **Data Understanding**

Text .

# **Importing Libraries**

# **Loading Data**

# **Inspecting the Data Set**

# **Checking types of data**

# **DATA PREPARATION**

# **Finding Unique Values**

# **Replacing Values (Syntax errors)**

# **Dropping Irrelevant Columns**

# **Renaming Columns**

# **Handling Missing Values**

# **Imputing Missing Values**

# **Handling Duplicate Rows**

# **Type Conversion**

# **Relationships**

# **Statistical Analysis**

# **Dealing Outliers**

# **Creating Dummies**

# **Scaling / Transformation**

# **MACHINE LEARNING - UNSUPERVISED LEARNING**

# **Principal Component Analysis (PCA)**

# **k-Means Clustering**

# **Creating examples to predict**

# **Preparing for analyses**

# **k-Means Clustering**

# **DATA DICTIONARY**

Text

# **References**