**CCT College Dublin**

**Assessment Cover Page**

| **Module Title:** | Strategically Thinking |
| --- | --- |
| **Assessment Title:** | Company Reviews |
| **Lecturer Name:** | James Garza |
| **Student Full Name:** | Group 2:  Ana Isabel Nieves Barcenas  Daniela Daia  Magdalene Ejiro Awaritefe  Rochana da Silva Matos  Karla Carolina Alvarado Minon |
| **Student Number:** | 2022455 - Ana Isabel  2017207 - Daniela  2022151 - Magdalene  2022175 - Rochana  2022461 - Karla Carolina |
| **Assessment Due Date:** | 4th December 2022 |
| **Date of Submission:** | 4Th December 2022 |

**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. |
| --- |

## THE IMPACT OF FEEDBACK: HOW REVIEWS LEAD TO CHANGE IN THE WORKPLACE

****

### by,

2022455 - Ana Isabel

2017207 - Daniela

2022151 - Magdalene

2022175 - Rochana

2022461 - Karla Carolina

**Higher Diploma in Science in Data Analytics for Business Strategic Thinking**

James Garza

CCT College Dublin, Ireland.

**Abstract**

We were tasked with finding a topic that interested us as a team; from there, we come up with a hypothesis or question to which we would need to find the answers. After several meetings, we realised that company reviews were an interesting subject.

As it is known, ratings and reviews are vital for any product or service acquisition. While ratings have only been around for about two decades, it’s hard to imagine shopping, for example, without them. According to consumer research conducted on over 30,000 global shoppers, most (88%) of shoppers use reviews to discover, evaluate and buy products or services.

This study aims to show witch factors are important to an employee to choose a company by using the data set: “Company\_Reviews” available at [Kaggle](https://www.kaggle.com/datasets/vaghefi/company-reviews) and the code is on [Google Collab](https://colab.research.google.com/drive/1oXrTTIQ3pMluEEOsdDuZllAcKZvYiPCR?authuser=1#scrollTo=X5TCaSW486qS) by following the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology, going through the stages: Business Understanding, Data Understanding, Data Preparation, Modelling and Evaluation and Deployment.

*Key-words:* consumer research, *company reviews, CRISP-DM,*

**Table of Contents**

Introduction

[Business Understanding 4](#_heading=h.gjdgxs)

[Data Understanding 6](#_heading=h.30j0zll)

[Data Preparation 9](#_heading=h.1fob9te)

[Modelling 15](#_heading=h.3rdcrjn)

[Evaluation 25](#_heading=h.1pxezwc)

[Deployment 25](#_heading=h.49x2ik5)

[Reference List 26](#_heading=h.2p2csry)

**Table of Figures**

Figure 1. Shape and info of the dataset 6

Figure 2. Head of the dataset 7

Figure 3. 5 number summary (describe) 7

[Figure 4. Dropping and renaming columns. 9](#_heading=h.3znysh7)

[Figure 5. Filling missing values with interpolate method. 9](#_heading=h.2et92p0)

[Figure 6. Calculating outliers. 10](#_heading=h.tyjcwt)

[Figure 7. Creating a consolidated dataset for Dublin. 10](#_heading=h.3dy6vkm)

[Figure 8. Boxplots per Dublin area 11](#_heading=h.1t3h5sf)

Figure 9. Boxplot for consolidated Dublin dataset 12

[Figure 10. Histogram for the original dataset (per area) 13](#_heading=h.4d34og8)

[Figure 11. Histograms per Dublin area 13](#_heading=h.2s8eyo1)

[Figure 12. Line graph per Dublin area 14](#_heading=h.17dp8vu)

[Figure 13. Line graph for Dublin consolidated dataset 15](#_heading=h.26in1rg)

[Figure 14. Decomposed series for Dublin consolidated dataset 16](#_heading=h.lnxbz9)

[Figure 15. Dickey-Fuller test for Dublin consolidated dataset 16](#_heading=h.35nkun2)

[Figure 16. Fitting the best ARIMA model into the training set for Dublin consolidated](#_heading=h.1ksv4uv) [dataset 18](#_heading=h.1ksv4uv)

[Figure 17. Residual Diagnostic for Dublin consolidated dataset 18](#_heading=h.44sinio)

[Figure 18. House prices with forecast for Dublin consolidated dataset 19](#_heading=h.2jxsxqh)

[Figure 19. Difference between predicted and actual prices for Dublin consolidated](#_heading=h.z337ya) [dataset 20](#_heading=h.z337ya)

[Figure 20. Chi-squared test for Dublin consolidated dataset 20](#_heading=h.3j2qqm3)

[Figure 21. Line graph for Dublin 01 dataset 21](#_heading=h.1y810tw)

[Figure 22. Decomposed series for Dublin 01 dataset. 21](#_heading=h.4i7ojhp)

[Figure 23. Dickey-Fuller test for Dublin 01 dataset 22](#_heading=h.2xcytpi)

[Figure 24. Fitting the best ARIMA model into the training set for Dublin 01 dataset. 22](#_heading=h.1ci93xb) [Figure 25. Residual Diagnostic for Dublin 01 dataset. 23](#_heading=h.3whwml4)

[Figure 26. House prices with forecast for Dublin 01 dataset 24](#_heading=h.2bn6wsx)

[Figure 27. Difference between predicted and actual prices for Dublin 01 dataset 24](#_heading=h.qsh70q)

[Figure 28. Chi-squared test for Dublin 01 dataset 24](#_heading=h.3as4poj)

# Introduction

# 

# 

# focar: We are predicting ratings using employee reviews factores.

we would like to know if the certain companies have better revenue,health and pension benefits, job security and safety, work life balance and will be rated higher

# 

# 

# 

# 

# 

# 

# 

# 

# 

# 

# 

# 

# 

# 

# 

# 

# 

# 

# 

# 

# 

# 

# 

# 

# 

# 

# 

# 

# 

# 

# 

# Business Understanding - Rochana:

he goal is to understand the business and its needs, meaning the objectives and requirements for this project.

**Hypothesis**

We are trying to understand and configure the data. Found in company reviews, that contribute to the rating of companies from best to worst.

We are trying to focus on these companies that rank best to worst and what are the catalysts that result in these rankings.

How do these reviews contribute to change in each company's workplace over a period of time?

Do bad reviews on companies affect their recruitment process?

**General Goal**

Statistical analysis rankings on what results in the most popular company.

What are the factors that lead to these ratings?

Are the factors general or specific to each company?

Research what benefits result in the outcome of satisfied employees in these companies.

**Successful outcome**

Successfully harvest the information and reasoning for satisfied employees in these companies

Ability to review data collected and make predictions based on what has been successful so far.

explore the review factors and whether these factors include salary, health and pension benefits, job security and safety, work life balance, rate the same in all companies.

We are predicting ratings using employee reviews factores.

we would like to know if the certain companies have better revenue,health and pension benefits, job security and safety, work life balance and will be rated higher

# Data Understanding - Carol

The information in this dataset is scraped from Indeed.com website containing information about companies and their employees' ratings and happiness, location, revenue, salaries and a lot of other useful information.

# Data Preparation - Isabel [2022455@student.cct.ie](mailto:2022455@student.cct.ie)

# Modelling - Magdalene [2022151@student.cct.ie](mailto:2022151@student.cct.ie)

# Evaluation - all

# Deployment - all

# Reference List - all

Brockwell, P.J. and Davis, R.A. (1987). *Time Series Theory and Methods*. Springer.

Brownlee, J. (2020). *Data Preparation for Machine Learning: Data Cleaning, Feature Selection and Data Transforms in Python*. Jason Brownlee.

Chatfield, C. and Xing, H. (2019). *The Analysis of Time Series: an Introduction with*

### *R*. 7th ed. CRC Press.

Kirchgssner, G. (2014). *Introduction to Modern Time Series Analysis*. Springer.

Ladley, J. (2016). *Mastering and Managing Data Understanding*. [online] CIO. Available at: https:/[/www.cio.co](http://www.cio.com/article/238649/mastering-and-managing-data-)m[/article/238649/mastering-and-managing-data-](http://www.cio.com/article/238649/mastering-and-managing-data-) understanding.html [Accessed 19 May 2022].

McKinney, W. (2017). *Python for Data Analysis*. 2nd ed. Beijing: O’Reilly Inc.

The pandas development team (2022). *pandas: Python Data Analysis Library*. [online] pandas. Available at: https://pandas.pydata.org/ [Acc