|  |  |  |
| --- | --- | --- |
| A picture containing text  Description automatically generated | **MSc Project Report** | |
| Student Name: Kirankumar Chaudhary | Matriculation Number: 2120051 | |
| Supervisor: Tiffany Young | Second Marker: Rob Lothian | |
| Course: MSc Data Science | | |
| Project Title: Energy Consumption Analysis and Optimisation for Scottish Councils | | |
| Start Date: 07/02/2023 | Submission Date: 25/04/2023 | |
| **CONSENT**  I agree  That the University shall be entitled to use any results, materials or other outcomes arising from my project work for the purposes of non-commercial teaching and research, including collaboration.  **DECLARATION**  **I confirm:**   * **That the work contained in this document has been composed solely by myself and that I have not made use of any unauthorised assistance.** * **That the work has not been accepted in any previous application for a degree.** * **All sources of information have been specifically acknowledged and all verbatim extracts are distinguished by quotation marks.** | | |
| Student Signature: | | Date Signed: Click or tap to enter a date. |

**Acknowledgement**

I would like to express my deepest gratitude to my project supervisor, Tiffany Young, for her unwavering support and guidance throughout the entirety of this project. Her expertise, patience, and dedication have been invaluable in helping me navigate the complexities of my research and address any doubts or questions that arose along the way. Thank you, Tiffany, for your unwavering support, and guidance.

# 

**Abstract**

I will write abstract here.

**Contents**

Acknowledgement ..…………………………………………………………………………………………………………… 2

Abstract .……………………………………………………………………………………………………………………………. 3

1 Introduction …………………………………………………………………………………………………………………… 4

1.1 Background ………………………………………………………………………………………………………

1.2 Motivation ………………………………………………………………………………………………………

1.3 Content of the rest of the report …………………………………………………………………

2 Literature Review ………………………………………………………………………………………………………

3 Project Specification………………………………………………………………………………………………………

3.1 Aim ……………………………………………………………………………………………

3.2 Objectives ……………………………………………………………………………………………

3.3 Functional and Non-Functional Requirements ………………………………………………

3.4 Methodology ……………………………………………………………………………………………

3.5 Project Plan ……………………………………………………………………………………………

3.6 Review of legal, ethical, social, professional and environmental issues …

3.7 Risks and safety ……………………………………………………………………………………………

4 Design ……………………………………………………………………………………………

5 Implementation ……………………………………………………………………………………………

6 Evaluation of work ……………………………………………………………………………………………

7 Conclusions and future works ……………………………………………………………………………………………

References ……………………………………………………………………………………………

Appendices ……………………………………………………………………………………………

**List of Figures**

**1 Introduction**

Some text comes here

**1.1 Background**

Energy consumption patterns are influenced by various factors, such as economic growth, population, climate, and technological advancements. Understanding these patterns is crucial for optimizing energy use and promoting sustainability. In recent years, data-driven techniques have emerged as powerful tools for analysing and predicting energy consumption. Machine learning algorithms have demonstrated significant potential in developing accurate models for various energy-related applications.

In the context of the Scottish council area, there is a need for robust data analysis and predictive modelling to better understand energy consumption trends, support policymaking, and drive optimization efforts. This project aims to address this need by collecting and analysing energy consumption data from 2005 to 2020 and utilizing machine learning techniques to develop accurate predictive models.

The objectives of the project include data collection from open-source websites, business intelligence dashboard development for monitoring and analysis, exploratory data analysis for insights and pattern identification, machine learning model development for accurate energy consumption predictions, and deploying these models to optimize energy consumption. Through these efforts, this project seeks to contribute to the ongoing initiatives aimed at creating a more sustainable and efficient energy landscape in the Scottish council area.

**1.2 Motivation**

The urgency of addressing energy consumption and sustainability concerns has become increasingly evident, particularly as global energy demand continues to grow. Efficient energy management and decision-making play a critical role in reducing the environmental and economic impact of energy consumption. The Scottish council area, like many other regions, faces the challenge of optimizing energy use to conserve resources, and reduce greenhouse gas emissions.

Developing accurate predictive models for energy consumption is essential for better strategic planning, policymaking, and infrastructure investments. Analysing energy consumption data from 2005 to 2020 can uncover valuable insights into the factors affecting energy consumption trends and how they have evolved over time. This knowledge can drive targeted improvements and optimization efforts.

The motivation behind this project is to leverage advanced data analytics and machine learning to provide a comprehensive understanding of energy consumption patterns and energy type is consumed in the Scottish council area. By conducting exploratory data analysis (EDA) and developing predictive models, we can facilitate informed decision-making by various stakeholders, including government authorities, utility companies, and end-users. This project will ultimately contribute to sustainable development, promote cost-effective energy management, and foster a greener future for Scotland and beyond.

**1.3 Content of the rest of the report.**

**2 Literature Review**

Energy consumption and its optimization have emerged as critical concerns in today's world, given the rising global energy demand, mounting environmental challenges, and the pressing need for sustainable energy systems. This project centres on collecting and analysing energy consumption data from the Scotland Council area spanning 2005 to 2020. The primary aim is to devise accurate predictive models for energy consumption and subsequently implement these models to optimize energy consumption patterns. This literature review delves into various aspects of energy consumption analysis, encompassing data collection and transformation, exploratory data analysis, business intelligence dashboards, predictive modelling, and optimization strategies. By thoroughly examining an array of studies and scholarly articles, this review elucidates the current state of research and best practices within the realm of energy consumption analysis and optimization.

**Are there any data sources that should be considered for a comprehensive understanding of energy consumption patterns in the Scotland Council area?**

The foundation of any successful energy consumption analysis and modelling lies in the accuracy and comprehensiveness of the data utilized. Numerous sources, including governmental and non-governmental organizations, provide open-source energy consumption data that can be leveraged for research and policy development. Notably, two key sources have emerged as crucial in the context of the Scotland Council area: the UK Department for Business, Energy & Industrial Strategy (BEIS) [1]and the Scottish Government [2]. These sources offer a wealth of data spanning various dimensions of energy consumption, providing researchers with ample opportunity to explore and analyse trends.

By accessing data from Scotland Government, one can gain insights into energy consumption patterns across different sectors, such as domestic, commercial, and public sector, as well as observe variations in usage within specific geographic locations spanning 2005 to 2020.

**Is there a dashboard available for the energy consumption of Scottish council areas from 2005 to 2020, and has exploratory data analysis been conducted on this energy consumption data**?

I could not find any specific dashboard or publication addressing energy consumption in Scottish council areas from 2005 to 2020. However, the UK Department of Energy & Climate Change [3] has published factsheet named as "Sub-national total final energy consumption statistics," which provides energy consumption exploratory data analysis (EDA) for sub-national regions in the UK spanning from 2005 to 2013, including Scotland.

Additionally, one can refer to the Scottish government's [4] official statistics on energy consumption. These sources provide comprehensive information on Scotland's energy consumption patterns and trends, but they not specifically focus on the council areas or cover the exact period from 2005 to 2020.

Given the available sources, I need to conduct your EDA and create a customized dashboard to analyse and visualize energy consumption trends in Scottish council areas during the specified period i.e., from 2005 to 2013.

References:

[3] **UK Department of Energy & Climate Change,** 2012. *Sub-national total final energy consumption statistics: factsheet*. [Online]. Department of Energy & Climate Change. Available from: <https://www.gov.uk/government/statistics/sub-national-total-final-energy-consumption-statistics-2010-factsheet> [Accessed Date]

[4] **Scottish Government**, n.d. *Energy statistics*. [Online]. Energy and Climate Change Directorate. Available from: <https://www.gov.scot/collections/energy-statistics/>

**What are the key factors that influence the choice between Power BI and Tableau for building dashboard for energy consumption data?**

Power BI and Tableau are both powerful data visualization and business intelligence tools. They have their unique strengths and limitations, which makes them suitable for different use cases. Power BI and Tableau offer various features for creating visually appealing and interactive dashboards. Tableau is known for its advanced visualization capabilities, while Power BI is appreciated for its ease of use, integration with Microsoft products, and lower cost (Biswal 2023).

One advantage of Power BI over Tableau is its seamless integration with other Microsoft products, such as Excel, SharePoint, and Azure. This makes it easier for users already familiar with the Microsoft ecosystem to adopt Power BI for creating dashboards. Furthermore, Power BI provides a more consistent user experience across different platforms and devices (Biswal 2023).

Power BI also has a more intuitive and user-friendly interface, making it easier for non-technical users to create dashboards without advanced coding or data analysis skills. The learning curve for Power BI is considered less steep compared to Tableau, which can make it more accessible for a wider audience. Another aspect where Power BI shines is in terms of cost-effectiveness. Power BI offers a more affordable pricing structure compared to Tableau, making it a better option for small businesses or organizations with limited budgets (Biswal 2023).

In conclusion, after considering factors such as seamlessness, my familiarity with the tool, ease of use, and cost-effectiveness, I have decided that Power BI is a better choice for me to create dashboards for this project.

References:

[5] Biswal A., 2023. *Power BI Vs Tableau: Difference and Comparison.* [Online]. Simplilearn. Available from: <https://www.simplilearn.com/tutorials/power-bi-tutorial/power-bi-vs-tableau> [ Accessed Date]

**Which programming language and software is best suited for performing exploratory data analysis (EDA)?**

When performing exploratory data analysis (EDA) in energy consumption data for the Scotland Council, there are several programming languages and software options available. Two popular options are Python and R, which are both widely used in data analysis tasks.

Python is widely used for EDA due to its versatility, readability, and the availability of a vast ecosystem of libraries specifically designed for data analysis and visualization. Some of the most notable libraries include pandas, NumPy, and Matplotlib, which offer robust data manipulation and visualization capabilities, making the EDA process efficient and accessible (McKinney 2017).

Additionally, Python's IPython and Jupyter Notebooks provide an interactive computing environment, enabling users to write, run, and visualize code all in one place, which is particularly useful for EDA. In "Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython," Wes McKinney provides a comprehensive guide for utilizing Python's capabilities to perform EDA. This book covers techniques and tools that enable users to efficiently clean, transform, and visualize data, thereby streamlining the process of discovering patterns and insights within the data (McKinney 2017).

R, on the other hand, is a programming language specifically designed for statistical computing and graphics (Ihaka and Gentleman 1996). It provides a wide range of statistical and graphical techniques, with packages like ggplot2 and dplyr that cater specifically to EDA tasks

Based on above findings, I have decided to use Python and Jupyter Notebook for EDA due to my familiarity with the language and the abundance of tools and libraries available for data analysis and visualization tasks.

Reference:

[6] McKinney, W., 2017. *Python for Data Analysis Data Wrangling with Pandas, NumPy, and IPython.* 2nd ed. Sebastopol, CA: O'Reilly Media, Inc.

[7] Ihaka, R. and Gentleman, R., 1996. R: A Language for Data Analysis and Graphics, *Journal of Computational and Graphical Statistics*, 5:3, pp. 299-314.

**How do different machine learning techniques, such as linear regression, support vector machines, artificial neural networks, and ensemble methods, compare in terms of predictive accuracy, complexity, and computational requirements when applied to energy consumption data?**

Different machine learning techniques have varying strengths and weaknesses when applied to energy consumption data, considering their predictive accuracy, complexity, and computational requirements.

Linear regression is a simple and interpretable model, making it easy to understand and implement. However, its simplicity limits its ability to capture complex relationships in the data, potentially leading to suboptimal predictive accuracy (James et al. 2013). Despite this limitation, linear regression may still perform well if the relationships in the energy consumption data are predominantly linear.

Support Vector Machines (SVM) are versatile models capable of handling both linear and nonlinear relationships by employing kernel functions. Compared to linear regression, SVMs may provide better predictive accuracy on complex datasets. However, they tend to be more computationally demanding, especially with large datasets, and their results may be more challenging to interpret (Cortes and Vapnik 1995).

Artificial Neural Networks (ANN) are powerful models inspired by biological neural networks. They can model complex relationships and nonlinearities in the data, often leading to improved predictive accuracy compared to simpler methods. However, ANNs come with higher complexity and increased computational requirements, and their results may lack interpretability (Goodfellow et al. 2016).

Ensemble methods, such as Random Forests, Gradient Boosting Machines (GBM), and eXtreme Gradient Boosting (XGBoost), combine multiple base models to improve predictive accuracy. These methods can handle complex data relationships and provide robust predictions, often outperforming single models. However, ensemble methods tend to be more computationally demanding and may require more extensive parameter tuning to achieve optimal performance (Hastie et al. 2009).

In summary, each machine learning technique has its advantages and disadvantages when applied to energy consumption data. The choice of the best method depends on the specific dataset, available computational resources, and the desired level of interpretability. To find the most suitable technique for a given problem, it is often helpful to perform cross-validation and compare the performance of different models on the data at hand.

References:

[8] James, G., Witten, D., Hastie, T., and Tibshirani, R., 2013. *An Introduction to Statistical Learning: with Applications in R*. New York: Springer.

[9] Cortes, C., and Vapnik, V., 1995. Support-vector networks. *Machine Learning*, 20(3), pp. 273-297.

[10] Goodfellow, I., Bengio, Y., and Courville, A., 2016. *Deep Learning*. Cambridge Massachusetts: MIT Press.

[11] Hastie, T., Tibshirani, R., and Friedman, J., 2009. *The Elements of Statistical Learning*. 2nd ed. New York: Springer.

**How do different evaluation metrics and criteria impact the selection of the most suitable predictive model for energy consumption analysis in varying contexts?**

The selection of the most suitable predictive model for energy consumption analysis in varying contexts depends on multiple factors, including the evaluation metrics and criteria used. Evaluation metrics and criteria, such as R-squared, mean squared error (MSE), mean absolute error (MAE), and root mean squared error (RMSE), have different implications for model performance and can impact the selection of the most suitable regression model.

R-squared is a measure of the proportion of the variance in the dependent variable that can be explained by the independent variables in the model. While a higher R-squared indicates better model fit, it does not necessarily imply the most accurate predictions, especially in the presence of overfitting (James et al., 2013) [10]. In some cases, models with lower R-squared may produce more accurate predictions for unseen data due to better generalization.

MSE and RMSE are measures of the average squared difference between the predicted values and the true values. While they effectively penalize large errors, they can be sensitive to outliers and may not accurately reflect the model's performance on most of the data (Hyndman & Koehler, 2006) [14].

MAE, on the other hand, measures the average absolute difference between the predicted and true values. It is less sensitive to outliers compared to MSE and RMSE and can provide a better indication of model performance when dealing with skewed data or extreme values (Willmott & Matsuura, 2005) [15].

The choice of evaluation metric and criteria depends on the specific context of the energy consumption analysis, such as the importance of accurately predicting extreme values or the distribution of the target variable. Selecting the most suitable model may require a trade-off between the various evaluation metrics, and the use of cross-validation and domain knowledge to guide this selection process (James et al., 2013) [10].

It is essential to critically examine the evaluation metrics and their implications for the specific context to select the most suitable predictive model for energy consumption analysis.

References:

[8] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning: with Applications in R. Springer.

[12] Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. International Journal of Forecasting, 22(4), 679-688.

[13] Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate Research, 30(1), 79-82.

**Practices should be considered when deploying machine learning model in local machine?**

The deployment of machine learning models is a critical phase in utilizing trained algorithms to make predictions and recommendations. It is crucial to integrate these models into production systems, where they can process new data and generate valuable insights [27]. When deploying machine learning models, several factors must be considered, including model performance, scalability, maintainability, and security [28]. In the context of energy consumption, the deployment of predictive models facilitates real-time monitoring, forecasting, and optimization. Research demonstrates the significance of deploying machine learning models in various energy-related applications such as demand response management [29], smart grid optimization [30], and building energy management systems [31]. The successful deployment of these models in production environments enables stakeholders to make data-driven decisions and implement energy-saving strategies.

Challenges in deploying machine learning models in real-world applications include ensuring model performance, scalability, maintainability, and security [28]. Researchers need to account for factors such as model interpretability, computational requirements, and model robustness to maintain optimal performance in production environments. Implementing best practices, such as continuous monitoring and updating of models, can address potential shifts in data distribution and evolving energy consumption patterns [32]. In the realm of energy consumption analysis, effective model deployment allows stakeholders to leverage predictive analytics' power, leading to informed decisions and optimized energy consumption.

**In conclusion**, the literature review has revealed several gaps in the existing research on energy consumption analysis, particularly with respect to the exploration and modelling of energy consumption data in the context of Scottish council areas from 2005 to 2020. The available sources and publications do not focus specifically on the council areas or cover the exact period specified for this project.

To address these gaps, this project will involve conducting exploratory data analysis (EDA) on energy consumption data in Scottish council areas during the specified period using Python and Jupyter Notebook. This choice is driven by the language's versatility, readability, and the abundance of tools and libraries available for data analysis and visualization tasks.

Additionally, Power BI has been chosen as the preferred tool for creating interactive dashboards and visualizations due to its ease of use, seamless integration with other Microsoft products, and cost-effectiveness. This will enable the integration of the generated dashboard into a PowerPoint presentation for easy reporting and presentation.

To select the most suitable predictive model for energy consumption analysis, multiple machine learning techniques will be explored and compared, including linear regression, support vector machines, artificial neural networks, and ensemble methods. The choice of the best method will be guided by cross-validation and domain knowledge, considering various evaluation metrics and criteria, such as R-squared, mean squared error, mean absolute error, and root mean squared error.

This project aims to fill the identified gaps in the literature by providing a comprehensive analysis of energy consumption data in Scottish council areas and developing an accurate and robust predictive model for future energy consumption. By critically examining the evaluation metrics, the context, and the existing research, this project will contribute valuable insights to the field of energy consumption analysis and modelling.

**3 Project Specification**

In this chapter

**3.1 Aim**

The primary objective of this project is to gather comprehensive energy consumption data from Scottish council areas spanning the years 2005 to 2020. By utilizing Power BI, the goal is to develop an informative and interactive dashboard to effectively visualize this data. Subsequently, Python and Jupyter Notebook will be employed for conducting an in-depth exploratory data analysis (EDA) to uncover valuable insights, patterns, and correlations within the dataset. Lastly, create, refine, and deploy highly accurate predictive models to enable optimized energy consumption, contributing to more sustainable and efficient energy management within these council areas.

**3.2 Objectives**

Collect energy consumption data for Scottish council area spanning over 2005-2020

from open-source websites.

Transform data according to needs.

Develop business intelligence dashboards and reports to effectively visualize the data.

Perform exploratory data analysis to gain insights into the data, identify

patterns, correlations, and anomalies.

Develop machine learning models to predict energy consumption accurately.

Deploy the machine learning models and the algorithms for optimization of energy

consumption

**3.3 Functional and Non-Functional requirements.**  
For this project on gathering energy consumption data for Scottish council areas, creating a dashboard, conducting EDA, and developing accurate predictive models, the functional and non-functional requirements can be described as follows:

**3.3.1 Functional Requirements:**

Data Collection: Collect comprehensive energy consumption data for Scottish council areas from 2005 to 2020 from reliable sources such as government websites or public databases.

Data Processing: Clean, pre-process, and transform the raw data into a structured format suitable for analysis and visualization.

Dashboard Creation: Develop an interactive and informative Power BI dashboard to visualize the energy consumption data, enabling users to easily explore and understand patterns and trends.

Exploratory Data Analysis: Perform in-depth EDA using Python and Jupyter Notebook to uncover insights, patterns, correlations, and anomalies within the energy consumption data.

Model Development: Create accurate predictive models for energy consumption based on the analyzed data, using machine learning algorithms and techniques.

Model Evaluation: Assess the performance of the developed models using suitable evaluation metrics and cross-validation techniques.

Deployment: Deploy the predictive models for practical use to optimize energy consumption and contribute to efficient energy management.

**3.3.2 Non-Functional Requirements:**

Usability: The dashboard and models should be user-friendly, with clear, intuitive, and easy-to-understand interfaces and visualizations.

Scalability: The developed models and tools should be able to handle increasing volumes of data as new information on energy consumption becomes available.

Performance: The models, EDA, and dashboard should have fast response times and low latency to provide a seamless user experience.

Maintainability: The code and tools used for the project should be well-structured, modular, and easy to maintain, allowing for future updates and improvements.

Security: Any sensitive data involved in the project should be properly protected, ensuring the privacy and confidentiality of the information.

Reliability: The models, EDA, and dashboard should provide accurate, consistent, and reliable results, with rigorous testing and validation processes in place.

Documentation: Comprehensive documentation should be provided for all aspects of the project, detailing the methods, tools, and processes used, as well as instructions for future maintenance and updates.

**3.4 Methodology**

**3.5 Project Plan**

**3.6 Review of legal, ethical, social, professional, and environmental issues**

**3.7 Risks and safety**

**4 Design**

In this chapter

**5 Implementation**

In this chapter

**6 Evaluation of work**

In this chapter

**7 Conclusions and Future of work**

In this chapter

**Appendices**

**Referencing**

[1] UK Department for Business, Energy & Industrial Strategy (BEIS), 2013. *Digest of UK Energy Statistics (DUKES)*. [Online]. Department for Energy Security and Net Zero and Department for Business, Energy & Industrial Strategy. Available from: <https://www.gov.uk/government/collections/digest-of-uk-energy-statistics-dukes> [Accessed Date].

[2] Scottish Government, 2014, *Energy Consumption*. [Online]. Scottish Government. Available from: <https://www.gov.scot/collections/energy-statistics/> [Accessed Date].

[3] UK Department of Energy & Climate Change, 2012. *Sub-national total final energy consumption statistics: factsheet*. [Online]. Department of Energy & Climate Change. Available from: <https://www.gov.uk/government/statistics/sub-national-total-final-energy-consumption-statistics-2010-factsheet> [Accessed Date]

[4] Scottish Government, n.d. *Energy statistics*. [Online]. Energy and Climate Change Directorate. Available from: <https://www.gov.scot/collections/energy-statistics/>

[5] Biswal A., 2023. *Power BI Vs Tableau: Difference and Comparison.* [Online]. Simplilearn. Available from: <https://www.simplilearn.com/tutorials/power-bi-tutorial/power-bi-vs-tableau> [ Accessed Date]

[6] McKinney, W., 2017. *Python for Data Analysis Data Wrangling with Pandas, NumPy, and IPython.* 2nd ed. Sebastopol, CA: O'Reilly Media, Inc.

[7] Ihaka, R. and Gentleman, R., 1996. R: A Language for Data Analysis and Graphics, *Journal of Computational and Graphical Statistics*, 5:3, pp. 299-314.

[8] James, G., Witten, D., Hastie, T., and Tibshirani, R., 2013. *An Introduction to Statistical Learning: with Applications in R*. New York: Springer.

[9] Cortes, C., and Vapnik, V., 1995. Support-vector networks. *Machine Learning*, 20(3), pp. 273-297.

[10] Goodfellow, I., Bengio, Y., and Courville, A., 2016. *Deep Learning*. Cambridge Massachusetts: MIT Press.

[11] Hastie, T., Tibshirani, R., and Friedman, J., 2009. *The Elements of Statistical Learning*. 2nd ed. New York: Springer.

[12] Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. International Journal of Forecasting, 22(4), 679-688.

[13] Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate Research, 30(1), 79-82.