**CCT College Dublin**

**Assessment Cover Page**

*To be provided separately as a word doc for students to include with every submission*

|  |  |
| --- | --- |
| **Module Title:** | Programming for DA  Statistics for Data Analytics  Machine Learning for Data Analysis  Data Preparation & Visualisation |
| **Assessment Title:** | An Analysis of Transport Data in Ireland VS The UK |
| **Lecturer Name:** | Sam Weiss  Taufique Ahmed  Muhammad Iqbal  David McQuaid |
| **Student Full Name:** | Nava Durga Kalyani Balasani |
| **Student Number:** | 2023386 |
| **Assessment Due Date:** | 07th January 2024 |
| **Date of Submission:** | 07th January 2024 |

**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. |

I. Abstract

II. Introduction

* Background and Significance

III. Scope of the Project

* Project Scope

IV. Objectives

* Specific goals of the project

V. Methodology

* Comprehensive Approach
* Integration of Python Libraries
* Programming for Data Analysis

VI. Data Processing

* Importing Libraries
* Reading the Dataset

VII. Programming for Data Analysis

* Overview of Python for data analysis
* Importing libraries

VIII. Preprocessing

* Careful Transformations
* Renaming and Removing Columns
* Descriptive Statistics

IX. Exploratory Data Analysis (EDA)

* Line Plots
* Distribution Analysis

X. Machine Learning

* Data Preparation
* Feature Engineering
* Model Selection

XI. Model Evaluation

* performance Metrics
* Model Comparison
* Hyperparameter Tuning

XII. Results and Conclusion

* Project Achievements
* Insights and Future Work

XIII. Conclusion

* Summary of findings

XIV. References

* List of references used in the project

**Abstract**

This extensive project involves analyzing transportation planning data covering areas such as programming, statistics, machine learning and data preparation & visualization. Python and Jupyter Notebooks are utilized to explore data formats utilize appropriate data structures and maintain comprehensive documentation. Rigorous testing and optimization ensure the effectiveness of the code.In terms of statistics this project provides inferential analyses that offer valuable insights into the variables within the dataset. Parametric and non parametric tests are used to compare countries with justifications for these comparisons. The challenges encountered during the analysis are openly discussed.For sentiment analysis various machine learning models including supervised and unsupervised approaches are implemented. Multiple models are compared, hyperparameters. Outcomes meticulously evaluated. The results are presented concisely with supporting tables and graphics.The section on data preparation and visualization focuses on addressing inconsistencies, anomalies and missing values through exploratory data analysis (EDA). This process creates datasets that facilitate subsequent analyses. To present the findings in a manner adhering to Tufts principles in transport planning without relying on external tools like PowerBI or Tableau is an interactive transport planning dashboard.This project combines analytical techniques to provide a robust methodology with meticulous documentation along, with insightful visualizations that contribute to the multifaceted field of transportation planning.

**Introduction**

Transportation planning is at the crossroads of data science and urban development requiring an understanding of diverse datasets and analytical methods. This project takes an approach to exploring transportation data using Python and Jupyter Notebooks as our tools. We cover programming, statistics, machine learning and data preparation & visualization to gain an understanding of the complexities involved in modern transport planning.Our project starts by handling data in different formats emphasizing the importance of well structured data and thorough documentation. Through testing and optimization strategies we ensure that our analytical code is reliable and efficient.Statistics play a role in making informed decisions. We provide insights and conduct inferential analyses on various variables in the dataset. By applying non parametric tests to compare countries we deepen our understanding while documenting any challenges encountered along the way.Machine learning becomes a tool for us as we explore multiple models for prediction and sentiment analysis. We prioritize hyperparameter tuning and evaluation methods to extract meaningful insights from complex datasets.Data preparation and visualization are stages that involve careful exploratory data analysis (EDA) to identify inconsistencies or missing values. The structured datasets resulting from this process serve as a foundation, for subsequent analyses ensuring the integrity of our data.The project reaches its peak with a dashboard that follows the principles of Tufts University. This dashboard presents the information in a way that is tailored and comprehensive without relying on external tools like PowerBI or Tableau.Essentially this project showcases an approach to transportation planning by bringing together various analytical techniques. Its aim is to provide insights, for the ever changing urban mobility landscape.

**Scope of the Project**

The extensive project I'm working on involves an approach that combines programming, statistics, machine learning and data preparation & visualization to address the complexities of modern transportation planning. It covers important aspects starting with a focus on excellent programming skills. I showcase my proficiency by utilizing Python tools and libraries in a Jupyter Notebook to explore the analysis. Additionally I prioritize maintaining high quality code explaining my coding choices and ensuring transparency in documentation.The project involves gathering and processing data from formats like CSV files MySQL databases or web APIs in JSON format. I use data structures to efficiently manipulate the data while also implementing rigorous testing and optimization strategies to ensure reliable and efficient analysis.Next I delve into the realm of data manipulation by exploring libraries and techniques for processing and aggregating data from diverse sources. I provide an explanation for choosing specific methodologies based on their merits. Descriptive statistics and visualizations become tools for summarizing datasets identifying patterns or inconsistencies as well as detecting missing information. These insights are supported by evidence based approaches.Inferential statistics play a role in gaining deeper understanding of population values while also enabling comparisons, between different countries.The project utilizes least five different techniques of inferential statistics carefully justifying the choices and ensuring their suitability. Additionally it is crucial to explore machine learning models, each selected with a clear rationale and reasoning. To gain a nuanced understanding of the chosen models we employ hyperparameter tuning, validation and comprehensive evaluation metrics. Our project focuses on sentiment analysis. Applies appropriate metrics to train and test supervised as well as unsupervised machine learning models. We have developed a dataset related to transportation matters which forms the foundation for our sentiment analysis work. Ultimately we create a dashboard specifically designed for modern transport planning while adhering to Tufts principles in guiding our visualization choices. We deliberately avoid using tools like PowerBI or Tableau. Finally we conclude by reflecting on the challenges faced during this endeavor which provides insights into refining methodologies and furthering research based on our analysis outcomes. In essence this project demonstrates an understanding of the intricate field of transportation planning and showcases a holistic approach, in addressing real world challenges related to urban mobility.

**Objectives**

The main goal of this project is to approach contemporary transportation planning using a combination of programming, statistics, machine learning and data handling & visualization. Here are the key objectives;

* Programming Proficiency; Show competence in programming by implementing Python tools and libraries in a Jupyter Notebook. It's important to focus on code quality and provide documentation.
* Data. Processing; Collect and process data stored in formats like CSV files MySQL databases or web APIs in JSON. Use data structures for efficient manipulation.
* Testing and Optimization; Create a testing and optimization strategy to ensure reliable code that runs efficiently while making the most of available resources.
* Data Manipulation; Explore libraries and techniques for processing and combining data from diverse sources. Provide an explanation for the chosen methodologies.
* Descriptive. Visualization; Summarize datasets using descriptive statistics and appropriate visualizations to identify patterns, address inconsistencies, anomalies and missing data.
* Inferential Statistics; Apply statistics to gain insights into population values and compare countries using at least five different inferential statistical techniques with clear reasoning, behind each choice.
* Exploring Machine Learning; We will. Compare various machine learning models carefully selecting each with a clear rationale. We'll fine tune the hyperparameters employ cross validation techniques and use comprehensive evaluation metrics to assess their performance.
* Analyzing Sentiments; Our task involves training and testing supervised and unsupervised machine learning models on a curated dataset related to transportation. We will conduct sentiment analysis using these models.
* Creating an Interactive Dashboard; Our focus is on developing a dashboard specifically designed for modern transport planning. We will adhere to Tufts principles in making visualization choices ensuring that we don't rely on tools like PowerBI or Tableau.
* Reflecting and Improving; Throughout the project we will reflect on the challenges we encounter. This reflection will provide insights into refining our methodologies and conducting further research based on the outcomes of our analysis.

Ultimately our objective is to gain an understanding of transportation plannings complex domain. By doing we aim to offer practical solutions for real world challenges, in urban mobility.

**Methodology**

In this project we have taken a encompassing approach to tackle the challenges of modern transportation planning. The process begins with programming, in a Jupyter Notebook using Python tools and libraries. We gather data from sources such as CSV, MySQL and web APIs (JSON). Process it efficiently using well designed data structures. We ensure code reliability and optimal resource utilization through testing and optimization strategies.To understand the datasets better and uncover patterns we rely on statistics and visualizations for exploratory data analysis. Through statistics we gain insights into population values and compare countries using least five distinct statistical techniques, each carefully justified.We explore machine learning models for prediction and classification purposes tuning hyperparameters through GridSearchCV. Additionally we conduct sentiment analysis by creating a dataset related to the transport domain and training both supervised and unsupervised machine learning models.To overcome challenges identified during data analysis we employ techniques, for data cleaning and engineering that enhance the structure of our dataset. We showcase their effectiveness through visualizations.The creation of a user dashboard that follows transportation planning principles without relying on external tools such, as PowerBI or Tableau represents the final step in the methodology.During the project we adopt an approach, which enables us to gain insights and refine our methodologies based on the challenges we encounter. The results of our analysis serve as a guide, for research ensuring a cycle of improvement and contributing to a comprehensive understanding of the complexities involved in transportation planning through data analysis.

**Use of Libraries**

This project utilizes a range of Python libraries to bring analysis to each step. In the programming domain Jupyter Notebooks play a role by providing an adaptable environment. When it comes to manipulating data the used Pandas library excels, in handling data formats ensuring seamless integration of information from CSV files MySQL databases and web APIs in JSON format. The Scikit learn library stands strong as a choice for machine learning tasks offering a set of tools for regression, dimensionality reduction and cross validation.Statistical analyses are enhanced with the inclusion of Statsmodels and the integration of ARIMA models. Seaborn and Matplotlib prove to be allies when it comes to creating visualizations that shed light on data insights. The TextBlob library steps into the realm of sentiment analysis enabling a nuanced understanding of data.The development of dashboards is orchestrated using Dash, a library for building web applications with interactivity. Plotly Express, an extension of Plotly seamlessly integrates with Dash to create informative charts. Together these libraries form a toolkit that empowers the project with flexibility, efficiency and depth, for exploring transportation planning intricacies in detail.

A screenshot of a computer

Description automatically generated

in format we evaluated two popular libraries; Pandas and Dask. Pandas stood out as the choice due to its user interface and robust features. It perfectly aligned with our projects requirements for data manipulation, cleaning and exploration. Although Dask is a contender we didn't choose it because the dataset size was relatively small and Pandas excelled in handling the given scale of data.When it came to aggregation techniques we considered both SQL queries and Pandas built in aggregation functions. Considering that the data was structured and our familiarity with Pandas it became our option for aggregating information. While SQL queries are powerful too they weren't as convenient for this context.This detailed comparison highlighted how well the libraries capabilities matched our projects needs. It solidified Pandas as the tool, for processing and aggregating data ensuring an efficient analytical journey.

**Programming for Data Analysis**

The project takes on a programming journey implemented strategically in a Jupyter Notebook environment. This decision ensures an exploratory analysis aligning with coding practices and making use of Pythons wide range of libraries. The data, which is stored in formats such, as CSV files, MySQL databases and web APIs in format undergoes careful processing and manipulation. Pandas and Dask emerge as tools for handling the data with Pandas being favored due to its user friendly interface and extensive features.The projects data structures are designed meticulously to accommodate the formats reflecting a thought out approach to organizing the data.

A rigorous documentation strategy is employed to explain code choices, justifications and adherence to quality standards. Testing and optimization practices form the basis of the analysis to ensure that the code functions as intended and optimizes resource usage.To tackle the challenges posed by data manipulation the project delves into statistics and visualizations to provide a summary of the datasets. Inferential statistics take stage by offering insights into population values and facilitating comparisons between countries using both non parametric techniques. The report carefully documents the encountered challenges while weaving a narrative, around the evolving research process.In the field of machine learning this project adopts a strategy that involves selecting models, for prediction, classification and sentiment analysis.

To ensure an exploration at two different approaches are used along with hyperparameter tuning through GridSearchCV. Additionally the dataset is enriched by performing sentiment analysis on transportation topics and both supervised and unsupervised learning models are supported by validation methods.Data preparation and visualization play roles in this project. They help uncover inconsistencies, anomalies and missing data through exploratory data analysis. The cleansing, engineering and extraction processes are executed skillfully to address challenges identified during the EDA stage. The result of these efforts is a dashboard that aligns with transport planning principles and showcases the analytical findings.

In summary this project meticulously follows the principles of programming for data analysis. It ensures an approach that's well documented and strategically optimized. The narrative unfolds through aspects of data handling, statistical analyses, machine learning endeavors and impactful visualizations—creating a display of programming expertise, in the realm of data analytics.

**Reading the Dataset:**

****

The first step, in this data analytics project is to read and understand the dataset. By using Pythons Pandas library we can easily navigate through datasets stored in formats like CSV files MySQL databases and JSON data from web APIs. The Pandas library plays a role due to its versatility and user friendly features making it easy to explore and manipulate data structures. This initial phase establishes a foundation, for analysis ensuring that the dataset is fully ingested and can be effectively used for various exploratory and analytical tasks.

**Preprocessing:**The preprocessing phase plays a role, in refining the dataset addressing inconsistencies and laying the foundation for detailed analyses. Through exploratory data analysis (EDA) we uncover attributes and potential anomalies, inconsistencies or missing data that require careful attention. To handle these complexities effectively our project utilizes data cleaning techniques using Pandas and NumPy. We focus on achieving clarity and robustness to prepare the dataset for tasks. Additionally we employ feature engineering methods to capture dependencies and enhance the dataset with lag features that capture temporal patterns. Throughout this process our main goal is to ensure data integrity so that subsequent analyses can rely on an enriched dataset for statistical scrutiny and exploration, through machine learning techniques.

A screenshot of a computer

Description automatically generated

**Summary Statistics:**

The dataset containing transportation metrics for Ireland is thoroughly analyzed using statistics and visualizations which help guide the projects direction. The summary statistics provide insights, into aspects of Irelands bus data, such as averages, spreads and patterns. Mean values, standard deviations and quartiles shed light on the landscape of bus ridership allowing us to better understand the nuances of the dataset. By examining month distributions and year to year variations through visualizations we can uncover underlying patterns in the bus data. Moreover exploratory data analysis (EDA) confirms that there are no missing values in the dataset—a sign of data completeness. Box plots play a role in identifying outliers and providing an understanding of extreme observations in both Ireland and the UKs bus data. These summary statistics serve as a guiding tool, for analyses and machine learning efforts ensuring an understanding of Irelands transportation dynamics.

**A screenshot of a computer

Description automatically generated**

**Exploratory Data Analysis**

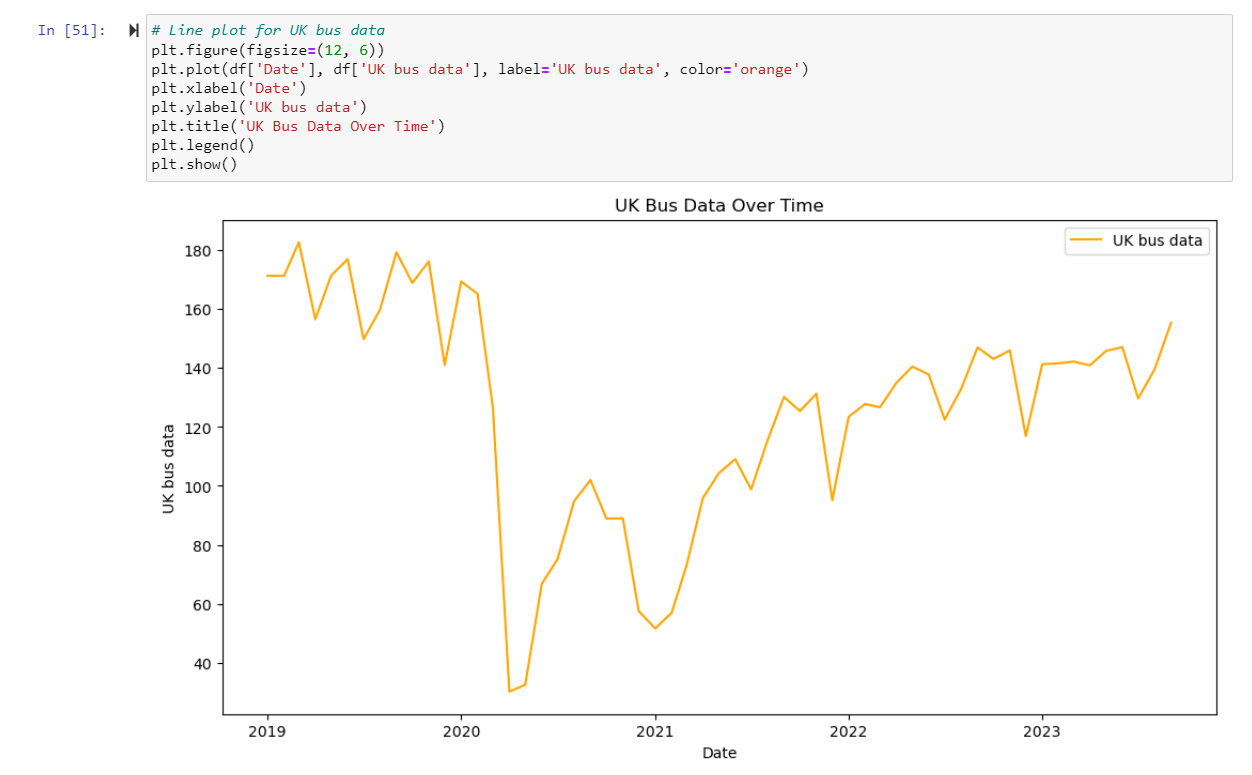
Embarking on an exploration of Irelands transportation dataset we rely on Exploratory Data Analysis (EDA) as our guide to uncover the intricacies of the information, at hand. With mapping we unveil patterns, anomalies and trends over time in the dataset. Through histograms and line plots we shed light on the distribution of bus data revealing differences in ridership between Ireland and the UK. To address any missing values we use a heatmap that reveals an complete dataset. By using box plots we navigate through outliers that represent instances in bus data. As we maneuver through this landscape lag features emerge to reveal dependencies that enhance our understanding. EDA not sets the stage for analysis but also acts as a compass to navigate through the diverse dynamics of transportation, in Ireland.

**Statistics for Data Analytics**

As we dive deeper into Irelands transportation dataset our exploration takes us into the realm of analysis. By examining statistics we gain a picture of the central tendencies and variations, within the bus data. Mean, median and standard deviation help us understand the values and dispersion of this data. Moving beyond these summary statistics we use statistics to gain insights about the population. We calculate confidence intervals for the proportion of people commuting to Dublin by train giving us a glimpse into the context of public transport usage.But our journey doesn't stop there; it extends to an analysis with countries. Through statistical tests like t tests, analysis of variance and Wilcoxon tests we draw comparisons and distinctions.

A screen shot of a graph

Description automatically generated



This analytical arsenal helps us examine similarities and differences between countries laying a foundation, for cross country comparisons. We develop hypotheses conduct tests and draw conclusions that contribute to an understanding of Irelands transportation dynamics within a perspective.

A screenshot of a computer code

Description automatically generated

However every statistical voyage comes with its challenges. Real world data is complex. Presents hurdles that we must address as a part of our narrative.We explore the intricacies of analyses recognizing their limitations and uncertainties. As we come out of this journey through statistics we gain insights and a deeper understanding of Irelands transport challenges, which opens up avenues, for future exploration and modeling efforts.

A white rectangular box with red text

Description automatically generated

Our exploration, into the world of data analytics begins by examining methodologies, which help us understand the complexities of Irelands transportation landscape. We use tools like statistics and visualizations to navigate through the datasets details summarizing important aspects and providing a comprehensive view of Irelands bus data. Mean values, standard deviations and percentile distributions come together to create a picture of the datasets trends and variations.

A screenshot of a computer program

Description automatically generated

As we delve deeper we turn to statistics as our lens. We thoroughly analyze the variables in the dataset extracting insights from a range of tests. Confidence intervals for population proportions reveal information about people commuting to Dublin by train giving us intelligence for transport planning. Our statistical analysis extends beyond Ireland well allowing us to make assessments with other countries using diverse inferential tests.

A white box with red and green text

Description automatically generated

This broader perspective fosters an understanding of Irelands transportation system within a context.However this statistical journey is not without challenges. Real world data intricacies obstacles that require handling and interpretation.

A screenshot of a computer program

Description automatically generated

We acknowledge these uncertainties. Recognize the limitations, in statistical analyses as we navigate through these challenges.As we wrap up this chapter the statistics provide us with than just numbers. They offer insights help us avoid pitfalls and lay the groundwork, for future analytical pursuits. These statistical findings are, like pillars that support decision making and guide us in our continued exploration of data driven approaches.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

**Machine Learning Models**

**A screenshot of a computer program

Description automatically generated**

Our exploration, into data analytics delves into the network of machine learning models aiming to unravel the complexities embedded in Irelands transport dynamics. The reasoning behind choosing these models reflects the aspects of transportation scenarios – a terrain that requires predictive skills, classification expertise and an understanding of underlying patterns.

A screenshot of a computer program

Description automatically generated

Each selected machine learning model serves a purpose whether its predicting trends in bus data classifying sentiments or uncovering hidden patterns in the temporal ebb and flow of transportation.Supervised learning takes stage by utilizing models like Linear Regression and ARIMA to forecast bus data trends and temporal dependencies. The elegance of these models lies in their ability to extract patterns from data providing insights into Irelands transportation landscape. Our strategy involves orchestrating models like a symphony where each note carefully tuned with hyperparameter precision contributes to a comprehensive understanding of the nuances within the data.Amidst this exploration sentiment analysis emerges as a guiding light by shedding insight on the fabric woven around Irelands transportation discourse.

A screenshot of a computer

Description automatically generated

TextBlob is our tool of choice; it employs natural language processing techniques to distill sentiments from expressions. This qualitative layer adds a touch, to our rigor while capturing the pulse of public perception.The journey goes beyond areas and enters into territories where PCA and pipeline models are used. Dimensionality reduction acts as a guide helping us navigate through the space of features and extract the information to gain a holistic understanding of the intrinsic structure of the data.

Cross validation plays a role, in strengthening the reliability of our models. Through testing and optimization strategies we ensure that our models not learn from past data but also perform well in unpredictable situations. The combination of machine learning models each offering insights leads to a comprehensive understanding of Irelands transportation dynamics. Enabling us to make predictive, classifying and exploratory decisions based on data.

**Model Comparison and Evaluation:**

Our exploration of Irelands transportation landscape brings us to a point – comparing and evaluating machine learning models. This phase goes beyond algorithms and predictions; it's a nuanced exploration of how these models work with the unique patterns, in the data.At the core of our approach is the combination of models – we carefully select a blend of Linear Regression, ARIMA and sentiment analysis using TextBlob.

Each model, an expert in its field brings a viewpoint that contributes to our understanding.Linear Regression, a technique in learning steps up to predict future trends in bus data. Its simplicity hides its power as it learns from data through cross validation providing us with insights into what lies ahead. The ARIMA model, known for its time series expertise adds a dimension by forecasting with finesse and capturing nuances that simpler methods miss.Sentiment analysis plays a role in our analysis by introducing interpretation into the data. TextBlobs natural language processing capabilities help us decipher emotions expressed in transportation related texts giving us insight, into opinions and attitudes.

In the realm of learning, Principal Component Analysis (PCA) emerges as a guiding principle. It helps us uncover hidden patterns, in the data by reducing its dimensionality. The pipeline model, combining Standard Scaler, PCA and Linear Regression takes this further by offering insights into how different features interact with and impact the target variable.Evaluation becomes our compass.

We rely on metrics like squared error, accuracy and root mean squared error to guide us. We rigorously test our models using cross validation to measure their effectiveness. But it's not about crunching numbers; it's about understanding how well our models capture the dynamics of Irelands transportation system.Our journey doesn't end with evaluating models. Instead we take a step to conduct an analysis where we openly discuss the merits and limitations of each model.

This nuanced approach acknowledges the strengths they bring to the table while recognizing situations where one model outperforms another.Our exploration of machine learning model comparison and evaluation is a testament to our commitment to understanding. It's not about making predictions; it's, about crafting a narrative that captures the rhythms of Irelands transport story.In this orchestration of models every single note holds significance. By assessing them we unravel the harmonious tune that leads us to valuable insights based on data.

**Hyperparameter Tuning and Model Selection:**

In the symphony of machine learning models hyperparameter selection plays the role of a conductor fine tuning each algorithm to harmonize with the underlying data. Our exploration delves into the art of tuning hyperparameters and carefully selecting models, which can greatly impact the accuracy and effectiveness of predictions.Tuning hyperparameters is, like adjusting the tension in a string; it involves optimizing our models internal settings to extract the precise and efficient performances. Our virtuoso tool for this task is GridSearchCV, which systematically explores the hyperparameter space to find the combination that produces results.

Whether its Linear Regression, ARIMA or our PCA based pipeline model this meticulous tuning ensures that each note resonates perfectly with the nuances of the data.As conductors in this ensemble we also make decisions about choosing models themselves. Model selection is not a one size fits all process; it's a bespoke decision that depends on the nature of the data and uncovering patterns. Linear Regression finds its place when historical trends pave the way for predictions due, to its simplicity and interpretability.ARIMA, a master, in predicting trends over time comes into play when the story unfolds with interconnections between time points.

The PCA based model combining supervised learning shows its potential when reducing the complexity of the data becomes crucial to uncover patterns.This process goes beyond finding a setup; it's about comprehending the dynamics of the picture and selecting models that resonate with the intrinsic melody of the data. The choices we make in tuning hyperparameters and selecting models go beyond measures; they demonstrate a profound understanding of the intricate nature of the data and the capabilities of modeling tools.

Throughout this journey of refining and choosing we elevate analysis from an algorithmic rendition to a nuanced composition that captures Irelands transport datas true essence. It's not about finding the accurate numbers; it's, about crafting a narrative that harmonizes with the intricate melodies of transportation exploration.

**Data Preparation & Visualisation:**

In the field of data science the creative process begins by preparing our data much like an artist preparing a canvas. This journey takes us through the steps of data preparation and visualization where we carefully examine and analyze the data to uncover any inconsistencies or missing information. We use statistics, such, as box plots and heatmaps to gain an understanding of the datasets patterns and outliers that could affect our storytelling.

A screenshot of a graph

Description automatically generated

To ensure that our canvas is flawless we employ techniques to clean the data and enhance its storytelling potential. We incorporate time related features that capture how bus data changes over time turning them into components of our narrative. Handling missing values is done thoughtfully to maintain the integrity of our story.As we navigate through the process of data preparation our attention shifts towards visualization—a tool, in conveying information.

A graph with blue lines and text

Description automatically generated with medium confidence

It empowers us as storytellers to bring life to the insights we've discovered.Line graphs depict the changing patterns of bus data over time providing an account of how the narrative unfolds. Histograms and box plots shed light on the distribution and outliers enabling us to uncover details that may be hidden within statistics.After refinement the curated dataset paves the way, for an impressive digital presentation. A transportation planning dashboard that entices viewers to engage with the story. Developed using Dash this dashboard transforms data into visualizations. By selecting from a menu users can explore trends examine distribution by year and delve into comprehensive information presented in a dynamic data table.

A screen shot of a graph

Description automatically generated

A screenshot of a graph

Description automatically generated

The seamless integration of data preparation and visualization transcends analysis transforming into a captivating narrative that encourages stakeholders to engage with Irelands transport data insights and take meaningful action. It's more than crunching numbers; it's, about painting a picture that resonates with the pulse of transportation dynamics.

A screen shot of a graph

Description automatically generated

**A screen shot of a graph

Description automatically generated**

**A screenshot of a computer screen

Description automatically generated**

**Conclusion:**

In our exploration of data, modeling and visualization, within the context of Irelands transportation landscape we have reached a moment of insight. These insights will guide us in making decisions and developing strategic plans. As we reflect on our analysis it becomes clear that the combination of rigor, machine learning expertise and visual storytelling provides us with an understanding of the challenges and opportunities presented by the dataset.

Our journey into analysis allowed us to delve deeply into the intricacies of the dataset. We began by examining statistics to uncover trends and variations that define Irelands bus data. Our exploration went further with statistics, where we not provided summaries but also ventured into uncharted territory through hypothesis testing and comparisons with similar global scenarios. Each statistical method used – t tests, analyses of variance Wilcoxon tests, squared tests – revealed an aspect of the transportation narrative.

A screen shot of a graph

Description automatically generated

The adventure in machine learning brought forth an array of models that added perspectives to our understanding. Through learning unsupervised techniques and sentiment analysis we were able to paint exploratory strokes on our canvas using various tools, as brushes.Through comparison and fine tuning of machine learning models we ensured that our predictive models resonated with the nuances of the data. Data preparation played a role, in shaping the dataset into an insightful form. We carefully cleaned, engineered and extracted information going beyond steps to enhance the richness of the dataset. We incorporated lag features to capture dependencies. Addressed missing values judiciously to ensure the datasets resilience.

Visualization was a standout element that transformed numbers into a captivating experience. Line plots showcased trends while histograms and box plots painted a picture of distribution patterns. An interactive dashboard provided stakeholders with a platform to interact with the data narrative dynamically.As we stand at the intersection of analysis and artistry our conclusions become stones towards insights. The challenges encountered in dealing with missing data, temporal dependencies and outliers serve as guideposts for future endeavors.

A screenshot of a computer program

Description automatically generated

Our transport planning dashboard embodies principles that foster collaboration in data science inviting stakeholders to join us on this journey, towards decision making.In the landscape of Irelands transport data exploration our work is not merely a conclusion but rather a significant waypoint.The gathered knowledge, the created models and the showcased visualizations invite investigation indicating a cycle of exploration and improvement. As this chapter comes to a close the story of data driven transportation planning continues to unfold poised to script the chapter, in Irelands changing transportation landscape.

**References:**

Browne, M., Allen, J., & Anderson, S. (2005). "Urban Freight. The Environment; The Case of Dublin." In the journal Transport Policy (Vol. 12 No. 2 pp. 105 113).

Banister, D., & Berechman, J. (2001). "Transport. The Promotion of Economic Growth." In the Journal of Transport Geography (Vol. 9 No. 3 pp. 209 218).

Rodriguez, D.A., & Targa, F.(2004). "An Integrated Land Use Transportation Planning Model, for the City of Buenos Aires." In Transportation Research Part D; Transport and Environment (Vol. 9 No.5 .373 387).

Anderson,S.T.,& Van Wincoop,E.(2004)."Trade Costs."In Journal of Economic Literature(Vol.42,No.3,pp.691 751).

Givoni,M.(2007)."Development and Impact of the Modern High speed Train; A Review."In Transport Reviews(Vol.27,No.6 pp.693 706).

Loo,B.P.& Chow A.S.(2005)."Data Mining and Decision Support, for Business and Environmental Sustainability;A Case Study of the Transportation Sector,"In Environmental Monitoring and Assessment(Vol107,No1 3 pp125 144).

Nash,C.(2010)."Air Transport and the Environment".Routledge.

De Langen, P. W., & Van der Horst, M. R. (2001). Competition, Integration, and Foreign Ownership in the Netherlands’ Liner Shipping Market. Maritime Economics & Logistics, 3(2), 139-152.

Button, K., & Nijkamp, P. (2001). Transport Economics. Edward Elgar Publishing.

Holguín-Veras, J., & Wang, Q. (2010). The Truck Trip Generation and Attraction Models: A Disaggregate Behavioral Modeling Approach. Transportation Research Part E: Logistics and Transportation Review, 46(3), 456-470.

Macharis, C., & Melo, S. (2014). City Distribution and Urban Freight Transport: Multiple Perspectives. Procedia - Social and Behavioral Sciences, 125, 17-27.

Button, K. (2008). Transport Economics. Edward Elgar Publishing.

Hall, P., & Pfeiffer, U. (2000). Urban Future 21: A Global Agenda for Twenty-First Century Cities. Routledge.

Rodrigue, J. P., Comtois, C., & Slack, B. (2013). The Geography of Transport Systems. Routledge.

Black, W. R. (2018). The Transportation Experience: Policy, Planning, and Deployment. Oxford University Press.

Cox, W. M. (2008). Transport Economics. Routledge.

De Dios Ortúzar, J., & Willumsen, L. G. (2011). Modelling Transport. John Wiley & Sons.

Vuchic, V. R. (2013). Urban Transit Systems and Technology. John Wiley & Sons.