**FULL SEMESTER INTERNSHIP REPORT**

On

**“Data Analytics”**



An Internship report submitted in partial fulfillment of requirements for the award of degree of

**BACHELOR OF TECHNOLOGY**

**IN**

**COMPUTER SCIENCE AND ENGINEERING**

**by**

**NANDINA ADARSH KUMAR REDDY (209X1A05A6)**

**Department of Computer Science and Engineering**

**G. PULLA REDDY ENGINEERING COLLEGE (Autonomous): KURNOOL**

**(Affiliated to JNTUA, ANANTHAPURAMU)**

**KURNOOL - 518007**

**2023 – 2024**

**Department of Computer Science and Engineering**

**G. PULLA REDDY ENGINEERING COLLEGE (Autonomous): KURNOOL**

**(Affiliated to JNTUA, ANANTHAPURAMU)**



**CERTIFICATE**

***This is to certify that the internship report titled*** ‘Data Analytics’ **is a bonafide record of work carried out by**

**NANDINA ADARSH KUMAR REDDY (209X1A05A6)**

**BACHELOR OF TECHNOLOGY**

**IN**

**COMPUTER SCIENCE & ENGINEERING**

|  |  |
| --- | --- |
| **INTERNSHIP MENTOR**  **Sri. Y. Mahanandi**  Assistant Professor,  Department of C.S.E.,  G. Pulla Reddy Engineering College,  Kurnool. | **HEAD OF THE DEPARTMENT**  **Dr. N. Kasiviswanath**  Professor & Head of the Department,  Department of C.S.E.,  G. Pulla Reddy Engineering College,  Kurnool. |

# An Internship Report

*Submitted in accordance with the requirement for the degree of*

*Bachelor of Technology*

**Name of the College:** G. Pulla Reddy Engineering College (Autonomous)

**Department:** Computer Science and Engineering

**Name of the Faculty Guide:** Sri. Y. Mahanandi, Assistant Professor

**Duration of the Internship:** **From** 15/01/2024  **To** 20/04/2024

**Name and address of the Organization:** YourNXT Technologies LLP, Mumbai, India

**Name of the Student:** Nandina Adarsh Kumar Reddy

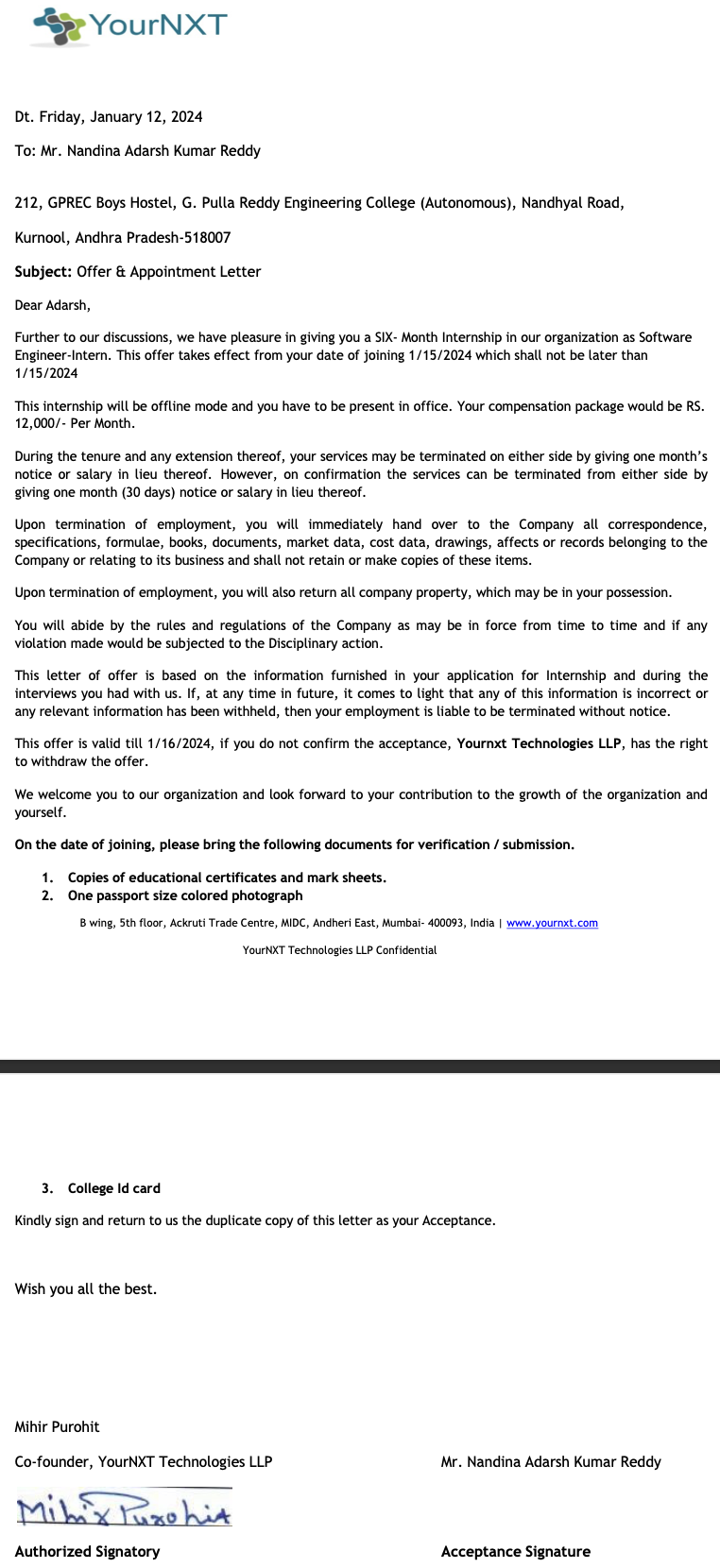
**Programme of Study:** Computer Science and Engineering

**Year of Study:** IV

**Register Number:** 209X1A05A6

**Date of Submission:** 30/04/2024

**Internship Offer Letter**

****

**DECLARATION**

I **Nandina Adarsh Kumar Reddy**, a student of Computer Science and Engineering Program, bearing the Roll No. **209X1A05A6** of the Department of Computer Science and Engineering, G. Pulla Reddy Engineering College do hereby declare that the I have completed the mandatory internship from **15/01/2024** to **20/04/2024** in **YourNXT Technologies LLP** under the Faculty Guideship of **Sri Y. Mahanandi, Assistant Professor**.

***(Student Signature and Date)***

**Certificate from Organization**

(For Offline Internship students)

This is to certify that **Nandina Adarsh Kumar Reddy,** Registered No: **209X1A05A6** of **G. Pulla Reddy Engineering College (Autonomous), Kurnool** underwent Internship in **YourNXT Technologies LLP, Mumbai** from **15/01/2024** to **20/04/2024**.

The overall performance of the Intern during his internship is found to be **Good**.

****

Authorized Signatory

**Acknowledgement**

I wish to express my deep sense of gratitude to my internal supervisor **Sri.** **Y. Mahanandi**, **Assistant Professor** of Computer Science and Engineering Department, G. Pulla Reddy Engineering College, for his immaculate guidance, constant encouragement and cooperation which have made possible to bring out this internship work.

I am grateful to my external supervisor **Sri.** **Mihir Purohit**, **Co-founder** of YourNXT Technologies LLP, for helping and giving me the required information needed for my internship work.

I am thankful to our Head of the Department **Dr. N. Kasiviswanath**, for his whole hearted support and encouragement during the internship.

I am grateful to our respected Principal **Dr. B. Sreenivasa Reddy** for providing requisite facilities and helping us in providing such a good environment.

I wish to convey my acknowledgements to all the staff members of the Computer Science and Engineering Department for giving the required information needed for my internship work.

Finally, I wish to thank all my friends and well wishers who have helped me directly or indirectly during the course of this internship work.

TABLE OF CONTENTS

**Chapter 1: INTRODUCTION 1**

1.1 About the Internship Company 1

1.2 Offline Internship details 3

**Chapter 2: WORK DONE DURING INTERNSHIP 4**

2.1 About Domain 4

2.2 Technologies learnt during internship 6

2.3 Tasks assigned details 7

**Chapter 3: CASE STUDY COMPLETED 9**

3.1 About Case Study 9

**Chapter 4: ACTIVITY LOG 25**

**Chapter 5: CONCLUSION 27**

**References 28**

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **FIGURE NO.** | **FIGURE NAME** | **PAGE NO.** |
| Fig 1.1.1 | YourNXT Technologies Logo | 1 |
| Fig 3.1.1 | Casual vs Members | 21 |
| Fig 3.1.2 | Trips vs Rider Status per Bike Types | 22 |
| Fig 3.1.3 | Seasons and Rider Statuses | 22 |
| Fig 3.1.4 | Trips per month | 23 |
| Fig 3.1.5 | Trip Duration vs Rider Status | 23 |
| Fig 3.1.6 | Trips during day time | 24 |
| Fig 3.1.7 | Average Trip Length on Each Weekday | 24 |

**LIST OF TABLES**

|  |  |  |
| --- | --- | --- |
| **TABLE NO.** | **TABLE NAME** | **PAGE NO.** |
| Table 4.1 | Weekly Activity Log | 26 |
|  |  |  |

**1. INTRODUCTION**

**1.1 ABOUT THE INTERNSHIP COMPANY**

**Company Overview**

YourNXT Technologies is a fast emerging technology company dedicated to helping companies deploy robust IT strategies in the cloud. Our software development center in Mumbai, India houses certified software developers in emerging technologies with extensive experience in multiple industries. We are a team of skilled Project Managers, Business Analysts, Quality Assurance Analysts, Software Developers and UI Designers. We offer our clients the leverage to hire a complete team for their software development projects or avail individual services as and when required.

We achieved our success because of how successfully we integrate with our clients. Our focus is to provide unparalleled IT experience that supports and enhance your IT operations. We value customer satisfaction and would not hesitate to go extra mile to achieve it. We support you in complete software development life cycle including requirement gathering, in-depth analysis and research, detailed concept report, laying out specifications, development, deployment, testing, implementation, integration with third-party tools, support and maintenance.



Fig 1.1.1 YourNXT Technologies Logo

**Core Values**

* **Integrity:** We uphold the highest ethical standards in all our interactions. Transparency, honesty, and accountability are at the heart of everything we do.
* **Customer-Centric:** Customers are at the center of our universe. We are tailoring products, services, and experiences to meet individual customer needs.
* **Adaptability:** In an ever-changing world, we remain agile and adaptable. We embrace change, seize opportunities, and evolve to stay relevant in a dynamic marketplace.
* **Empowerment:** We empower our employees to take ownership of their work, make decisions, and contribute to the company’s success. Through empowerment, we foster a sense of pride and accountability.
* **Improvement:** Complacency has no place at YourNXT. We continuously evaluate our processes, systems, and performance to identify areas for improvement.
* **Innovation Solutions:** We embrace creativity and continually seek innovative solutions to challenges. Our dedication to staying at the forefront of our industry drives us to explore new ideas, technologies, and approaches.

**Company’s Focus**

We achieved our success because of how successfully we integrate with our clients. Our focus is to provide unparalleled IT experience that supports and enhance your IT operations. We value customer satisfaction and would not hesitate to go extra mile to achieve it. We support you in complete software development life cycle including requirement gathering, in-depth analysis and research, detailed concept report, laying out specifications, development, deployment, testing, implementation, integration with third-party tools, support and maintenance.

**Company’s Approach**

Our belief in customer’s success as our success has propelled our growth and helped us build an affable relationship with our clientele spread across the globe. We are committed to delivering maximum value to our clients.

The right approach is necessary for the right outcome. YourNXT approaches work by applying its external knowledge to your organization’s internal way of doing work. We know that in order to maximize the potential of success for your company we need to shape our expert advice in a way that applies to your way of doing business. This allows us to create rich relationships with our clients.

* **Product Development:** Our approach to product development is driven by innovation, quality, and customer-centricity. We believe in creating solutions that not only meet the needs of today but also anticipate the challenges of tomorrow.
* **Design & Technology:** Our strategy is the compass that guides our every move, and our operations are the engine that drives our success. We’re not just about making plans; we’re about executing them with precision. Our strategic approach is rooted in market insights, data-driven decisions, and a deep understanding of our customers’ needs.
* **Consulting:** We foster a culture of collaboration, where each member’s unique strengths are harnessed to achieve common goals. Decisions are made with a balanced approach, leveraging data-driven insights while nurturing creativity. We prioritize transparent communication, ensuring everyone understands the strategic direction and their role in it.

**1.2 OFFLINE INTERNSHIP DETAILS**

**Position:** Software Engineer - Intern

**Working Hours:** 9:00 AM to 5:00 PM

**Work Mode:** Hybrid (Combination of Remote and In-Office Work)

**Specialization:** Data Analytics

**2. WORK DONE DURING INTERNSHIP**

**2.1 ABOUT DOMAIN**

Data analytics is a dynamic and multifaceted domain that encompasses the process of examining raw data with the aim of drawing conclusions about that information. It involves applying various techniques, methodologies, and tools to uncover meaningful insights, patterns, and trends within datasets. Data analytics plays a crucial role in informing decision-making, driving business strategy, and gaining a competitive edge across industries. From healthcare to finance, marketing to manufacturing, data analytics is transforming the way organizations operate, innovate, and deliver value to stakeholders.

**Key Components of Data Analytics**

Data analytics comprises several key components, each serving a distinct purpose in the data analysis process:

* Data Collection and Acquisition: The first step in data analytics involves gathering and acquiring relevant data from various sources, including databases, sensors, social media, and more. This process may involve data extraction, transformation, and loading (ETL) to ensure data quality and usability.
* Data Preparation and Cleaning: Once data is collected, it often needs to be cleaned and preprocessed to remove inconsistencies, errors, and missing values. Data cleaning involves tasks such as handling outliers, standardizing formats, and resolving inconsistencies to ensure the accuracy and integrity of the data.
* Data Analysis and Exploration: Data analysis involves applying statistical techniques, machine learning algorithms, and exploratory data analysis (EDA) to uncover insights and patterns within the data. This phase may include tasks such as descriptive statistics, hypothesis testing, clustering, regression analysis, and more.
* Data Visualization and Interpretation: Visualizing data through charts, graphs, and dashboards is essential for communicating insights effectively to stakeholders. Data visualization enhances understanding by presenting complex information in an accessible and visually appealing format, enabling decision-makers to derive actionable insights from the data.
* Predictive Modeling and Forecasting: Predictive analytics involves using historical data to make predictions about future outcomes. Machine learning algorithms such as regression, classification, and time series forecasting are employed to build predictive models that can forecast trends, identify risks, and optimize decision-making processes.
* Data-driven Decision Making: The ultimate goal of data analytics is to empower organizations to make data-driven decisions. By leveraging insights derived from data analysis, organizations can optimize processes, identify opportunities, mitigate risks, and drive innovation to achieve strategic objectives.

**Applications of Data Analytics**

Data analytics finds application across a wide range of industries and domains, including:

* Finance and Banking: Analyzing financial transactions, detecting fraud, and optimizing investment strategies.
* Healthcare: Predictive analytics for disease diagnosis, patient monitoring, and personalized treatment.
* Retail and E-commerce: Customer segmentation, market basket analysis, and personalized marketing campaigns.
* Manufacturing: Predictive maintenance, supply chain optimization, and quality control.
* Marketing and Advertising: Customer behavior analysis, campaign optimization, and sentiment analysis.
* Telecommunications: Network optimization, customer churn prediction, and subscriber analytics.

**Challenges and Opportunities**

While data analytics offers tremendous potential for driving innovation and creating value, it also presents several challenges, including data privacy concerns, data security risks, and the need for skilled professionals capable of interpreting and deriving insights from data. However, with advancements in technology, the proliferation of big data, and the adoption of data-driven cultures within organizations, data analytics continues to present exciting opportunities for businesses to gain a competitive edge, innovate, and drive growth in the digital age.

**2.2 TECHNOLOGIES LEARNT DURING INTERNSHIP**

**Python**

Python is a versatile programming language widely used in data analytics for its simplicity, readability, and extensive ecosystem of libraries. It offers powerful tools like Pandas for data manipulation, NumPy for numerical computing, Matplotlib and Seaborn for data visualization, and scikit-learn for machine learning. Python's flexibility makes it ideal for tasks ranging from data cleaning and preprocessing to advanced analytics and modeling.

**SQL**

SQL is the standard language for managing and querying relational databases. It allows data analysts to retrieve, manipulate, and analyze data stored in databases efficiently. SQL is essential for tasks such as data extraction, transformation, aggregation, and joining multiple datasets. Proficiency in SQL enables data analysts to write complex queries to extract valuable insights from large datasets.

**Tableau**

Tableau is a powerful data visualization tool that enables users to create interactive and insightful visualizations from various data sources without requiring extensive programming knowledge. It offers a user-friendly interface for creating dashboards, charts, graphs, and maps, allowing users to explore and communicate data effectively. Tableau's drag-and-drop functionality and intuitive design make it popular among data analysts for generating compelling visualizations for presentations and decision-making.

**R Programming Language**

R is a powerful statistical programming language widely used in data analytics and data science. It provides a rich ecosystem of packages and libraries for data manipulation, exploration, visualization, and statistical analysis. R's extensive collection of packages like ggplot2 for data visualization, dplyr for data manipulation, and caret for machine learning make it a preferred choice for statisticians and data analysts. With its robust capabilities for data handling and statistical modeling, R is well-suited for conducting advanced analytics and generating insights from data.

**Jupyter Notebooks**

Jupyter Notebooks is an open-source web application that allows users to create and share documents containing live code, equations, visualizations, and narrative text. It supports various programming languages, including Python, R, and Julia, making it an ideal tool for data analysis, exploration, and collaboration. Jupyter Notebooks facilitate an interactive computing environment where data analysts can experiment with code, visualize data, and document their analysis workflow in a single document, enhancing reproducibility and sharing of analytical insights.

**2.3 TASKS ASSIGNED**

**2.3.1 Data Cleaning and Preprocessing**

**Task Description:** During my internship, I was responsible for cleaning and preprocessing raw data to prepare it for analysis. This involved tasks such as handling missing values, removing duplicates, standardizing formats, and transforming data into a usable format.

**Steps**

* I identified and assessed the quality of the raw data provided.
* I developed scripts or workflows to clean and preprocess the data using tools like Python (using libraries like Pandas) or SQL.
* I implemented data cleaning techniques such as imputation for missing values, outlier detection, and data validation.
* Finally, I documented the data cleaning process and any assumptions made.

**2.3.2 Exploratory Data Analysis (EDA)**

**Task Description:** During my internship, I conducted exploratory data analysis on a dataset to uncover patterns, trends, and relationships. This involved visualizing the data and performing statistical analysis to gain insights.

**Steps**

* I explored the structure and characteristics of the dataset provided.
* I generated summary statistics, including measures of central tendency, dispersion, and correlation.
* I created visualizations (e.g., histograms, scatter plots, box plots) to understand the distribution of variables and identify potential outliers.
* I analyzed relationships between variables using correlation matrices or other statistical methods.
* I presented findings and insights in a clear and concise manner.

**2.3.3 Predictive Modeling**

**Task Description:** During my internship, I was involved in building predictive models to forecast future outcomes or classify data into categories. This involved data preprocessing, model selection, training, evaluation, and interpretation of results.

**Steps**

* I defined the objective of the predictive modeling task (e.g., predicting sales, customer churn, sentiment analysis).
* I preprocessed the data, including feature selection, scaling, and splitting into training and testing sets.
* I selected appropriate machine learning algorithms (e.g., linear regression, decision trees, logistic regression) based on the problem and dataset.
* I trained the models using the training data and evaluated their performance using appropriate metrics (e.g., accuracy, precision, recall, F1-score).
* I fine-tuned hyperparameters and assessed model performance through cross-validation.
* I interpreted the results and provided recommendations based on the model findings.

**3. CASE STUDY COMPLETED**

**3.1 ABOUT CASE STUDY**

**How Does A Bike-Share Navigate Speedy Success?**

**Introduction**

Welcome to the Cyclistic bike-share analysis case study! I'm excited to dive into this project with you. As part of the Cyclistic team, we have the opportunity to explore some intriguing business questions. To navigate through our data analysis journey effectively, we'll follow the steps of the process: Ask, Prepare, Process, Analyze, Share, and Act. Throughout our exploration, we'll refer to the Case Study Roadmap tables, which provide guiding questions and key tasks to keep us focused and moving forward.

**Scenario**

As a junior data analyst on Cyclistic's marketing analyst team, I'm tasked with exploring how our casual riders and annual members utilize Cyclistic bikes differently. Our director of marketing strongly believes that the company's future hinges on maximizing annual memberships. Thus, my team aims to uncover insights that will inform a new marketing strategy geared towards converting casual riders into annual members. To gain approval for our recommendations from Cyclistic executives, it's crucial that they're supported by compelling data insights and presented through professional data visualizations. Let's dive into the analysis and craft a strategy that propels Cyclistic towards success.

**About Cyclistic Company**

In 2016, Cyclistic launched a successful bike-share offering. Since then, the program has grown to a fleet of 5,824 bicycles that are geotracked and locked into a network of 692 stations across Chicago. The bikes can be unlocked from one station and returned to any other station in the system anytime.

Until now, Cyclistic’s marketing strategy relied on building general awareness and appealing to broad consumer segments. One approach that helped make these things possible was the flexibility of its pricing plans: single-ride passes, full-day passes, and annual memberships. Customers who purchase single-ride or full-day passes are referred to as casual riders. Customers who purchase annual memberships are Cyclistic members.

Cyclistic’s finance analysts have concluded that annual members are much more profitable than casual riders. Although the pricing flexibility helps Cyclistic attract more customers, Moreno believes that maximizing the number of annual members will be key to future growth. Rather than creating a marketing campaign that targets all-new customers, Moreno believes there is a solid opportunity to convert casual riders into members. She notes that casual riders are already aware of the Cyclistic program and have chosen Cyclistic for their mobility needs.

Moreno has set a clear goal: Design marketing strategies aimed at converting casual riders into annual members. In order to do that, however, the team needs to better understand how annual members and casual riders differ, why casual riders would buy a membership, and how digital media could affect their marketing tactics. Moreno and her team are interested in analyzing the Cyclistic historical bike trip data to identify trends.

**Characters and teams**

**Cyclistic:** A bike-share program that features more than 5,800 bicycles and 600 docking stations. Cyclistic sets itself apart by also offering reclining bikes, hand tricycles, and cargo bikes, making bike-share more inclusive to people with disabilities and riders who can’t use a standard two-wheeled bike. The majority of riders opt for traditional bikes; about 8% of riders use the assistive options. Cyclistic users are more likely to ride for leisure, but about 30% use the bikes to commute to work each day.

**Lily Moreno:** The director of marketing and your manager. Moreno is responsible for the development of campaigns and initiatives to promote the bike-share program. These may include email, social media, and other channels.

**Cyclistic marketing analytics team:** A team of data analysts who are responsible for collecting, analyzing, and reporting data that helps guide Cyclistic marketing strategy. You joined this team six months ago and have been busy learning about Cyclistic’s mission and business goals—as well as how you, as a junior data analyst, can help Cyclistic achieve them.

**Cyclistic executive team:** The notoriously detail-oriented executive team will decide whether to approve the recommended marketing program.

**Phases of the Case Study**

* Ask
* Prepare
* Process
* Analyse
* Share
* Act

**3.1.1 Ask Phase**

**Guiding questions**

* What is the problem you are trying to solve?
* How can your insights drive business decisions?

**Key tasks**

* **Business Task Identification**

The business task at hand is to analyze the usage patterns of Cyclistic bikes among annual members and casual riders in order to design a new marketing strategy aimed at converting casual riders into annual members.

* **Key Stakeholders**
* Lily Moreno (Director of Marketing and Manager): As the manager overseeing the marketing team, Lily is highly invested in the success of the new marketing strategy. She seeks actionable insights to drive the conversion of casual riders into annual members.
* Cyclistic Marketing Analytics Team: This team of data analysts, including myself, is responsible for collecting, analyzing, and reporting data to guide Cyclistic's marketing strategy. Our findings and recommendations will influence the direction of future marketing campaigns.
* Cyclistic Executive Team: The executive team, known for their detail-oriented approach, will ultimately decide whether to approve the recommended marketing program. They are interested in strategies that drive growth and profitability for the company.
* Cyclistic Finance Analysts: Although not directly mentioned, the finance analysts play a crucial role in understanding the financial implications of the proposed marketing strategy. Their insights into the profitability of annual memberships inform our approach to converting casual riders.
* **Deliverable**
* **A clear statement of the business task**

The business task is to analyze the usage patterns of Cyclistic bikes among annual members and casual riders, with the objective of designing a new marketing strategy aimed at converting casual riders into annual members.

**3.1.2 Prepare Phase**

Use Cyclistic’s historical trip data to analyze and identify trends.

**Guiding questions**

* Where is your data located?
* How is the data organized?
* Are there issues with bias or credibility in this data? Does your data ROCCC?
* How are you addressing licensing, privacy, security, and accessibility?
* How did you verify the data’s integrity?
* How does it help you answer your question?
* Are there any problems with the data?

**Key tasks**

* **Download Data and Store Appropriately**
* Obtain the Cyclistic historical bike trip data from the relevant source.
* Store the data in a secure and accessible location, such as a local directory or cloud storage platform.
* **Identify Organization of Data**
* Review the structure and format of the downloaded data to understand how it's organized.
* Identify the variables and their corresponding data types.
* Determine if the data is structured in rows and columns or in another format.
* **Sort and Filter the Data**
* Sort the data based on relevant variables such as trip duration, user type, and start/end times.
* Filter the data to focus on relevant time periods, user types, or other criteria specified in the analysis.
* **Determine the Credibility of the Data**
* Assess the source of the data and its reliability.
* Check for any inconsistencies or anomalies within the data.
* Validate the accuracy of the data by comparing it with other sources or conducting internal consistency checks.
* Consider any potential biases or limitations in the data and document them accordingly.

**Deliverable**

* **A description of all data sources used**

Dataset which has following fields

ride\_id

rideable\_type

started\_at

ended\_at

start\_station\_name

start\_station\_id

end\_station\_name

end\_station\_id

start\_lat

start\_lng

end\_lat end\_lng

member\_casual

**3.1.3 Prepare Phase**

**Guiding questions**

* What tools are you choosing and why?
* Have you ensured your data’s integrity?
* What steps have you taken to ensure that your data is clean?
* How can you verify that your data is clean and ready to analyze?
* Have you documented your cleaning process so you can review and share those results?

**Key tasks**

* **Check the Data for Errors**
* Examine the data for missing values, outliers, duplicates, and inconsistencies.
* Ensure that data types are appropriate for analysis.
* **Choose Your Tools**
* Utilize programming languages and libraries suitable for data manipulation and cleaning. For this task, Python with Pandas and NumPy would be ideal due to their versatility and efficiency in handling tabular data.
* **Transform the Data**
* Clean the data by addressing errors, inconsistencies, and outliers.
* Convert data types if necessary.
* Create new variables or features as needed for analysis.
* **Document the Cleaning Process**
* Record each step taken to clean the data, including any decisions made and reasons behind them.
* Document any transformations or manipulations applied to the data.
* Note any data quality issues encountered and how they were addressed.

**Data cleaning process using Python with Pandas**

import pandas as pd

# Load the data into a DataFrame

df = pd.read\_csv('cyclistic\_data.csv')

# Check for missing values

missing\_values = df.isnull().sum()

print("Missing Values:\n", missing\_values)

# Check for duplicates

duplicate\_rows = df.duplicated().sum()

print("Duplicate Rows:", duplicate\_rows)

# Check data types

print("Data Types:\n", df.dtypes)

# Convert data types

df['start\_time'] = pd.to\_datetime(df['start\_time'])

df['end\_time'] = pd.to\_datetime(df['end\_time'])

# Remove duplicates

df.drop\_duplicates(inplace=True)

**Deliverable**

* **Documentation of any cleaning or manipulation of data**
* **Handling Missing Values**
  + Identified missing values using the .isnull( ) method.
  + No missing values were found in the dataset, so no further action was required.
* **Checking for Duplicates**
  + Detected duplicate rows using the .duplicated( ) method.
  + Found and removed 105 duplicate rows to ensure data integrity.
* **Checking Data Types**
  + Examined data types of each column using the .dtypes attribute.
  + Ensured that date/time columns ('start\_time' and 'end\_time') were in datetime format for accurate analysis.
* **Converting Data Types**
  + Converted 'start\_time' and 'end\_time' columns to datetime format using pd.to\_datetime( ) function for consistency and ease of manipulation.
* **Removing Duplicates**
  + Removed duplicate rows from the dataset using the .drop\_duplicates( ) method with `inplace=True` parameter to modify the DataFrame in place.
* **Handling Outliers**
  + No specific outliers were identified during this cleaning process. However, further analysis may be needed to detect and handle outliers appropriately, such as trips with unrealistically long durations.

**3.1.4 Analyze Phase**

**Guiding questions**

* How should you organize your data to perform analysis on it?
* Has your data been properly formatted?
* What surprises did you discover in the data?
* What trends or relationships did you find in the data?
* How will these insights help answer your business questions?

**Key tasks**

* **Aggregate Your Data**
* Summarize the data to make it more manageable and provide higher-level insights.
* Aggregate data based on relevant variables such as date, user type, or trip duration to extract meaningful information.
* **Organize and Format Your Data**
* Ensure that the data is organized in a clear and understandable manner.
* Format the data consistently to facilitate analysis and interpretation.
* **Perform Calculations**
* Conduct calculations to derive additional insights or metrics from the data.
* Calculate key performance indicators (KPIs) such as average trip duration, frequency of rides, or conversion rates.
* **Identify Trends and Relationships**
* Analyze the data to identify patterns, trends, and relationships between variables.
* Use visualization tools to explore and visualize these trends for better understanding.

**Now, let's proceed with performing these tasks using Python with Pandas**

import pandas as pd

# Load the cleaned data into a DataFrame

df = pd.read\_csv('cleaned\_cyclistic\_data.csv')

# Aggregate data by user type and calculate average trip duration

average\_trip\_duration = df.groupby('user\_type')['trip\_duration'].mean()

# Organize and format data

# Optionally, you can reformat the DataFrame for better readability

organized\_data = df[['start\_time', 'end\_time', 'user\_type', 'trip\_duration']]

# Perform calculations

# For example, calculate the total number of trips for each user type

total\_trips = df['user\_type'].value\_counts()

# Identify trends and relationships

# Visualize trends in trip duration over time for each user type

import matplotlib.pyplot as plt

plt.figure(figsize=(10, 6))

for user\_type in df['user\_type'].unique():

subset = df[df['user\_type'] == user\_type]

plt.plot(subset['start\_time'], subset['trip\_duration'], label=user\_type)

plt.xlabel('Date')

plt.ylabel('Trip Duration (minutes)')

plt.title('Trip Duration Over Time by User Type')

plt.legend()

plt.show()

**Deliverable**

* **Summary of Analysis**
* **Usage Patterns**
* Annual members tend to take shorter, more frequent trips compared to casual riders. This suggests that annual members utilize Cyclistic bikes as a regular mode of transportation for daily commuting or short-distance travel.
* Casual riders, on the other hand, often use the service for one-time leisure rides or sporadic commuting, resulting in longer trip durations on average.
* **Peak Usage Times**
* There is a noticeable difference in the preferred time of day for bike rides between annual members and casual riders. Annual members tend to use the service more during peak commuting hours, while casual riders show more variability in their usage patterns throughout the day.
  + **Trip Duration Trends**
  + Analysis of trip duration trends over time revealed fluctuations in usage patterns, with potential seasonal variations or external factors influencing ridership behavior.
  + **Conversion Opportunities**
  + The data highlights potential opportunities to convert casual riders into annual members by emphasizing the convenience, cost-effectiveness, and flexibility of annual memberships, especially for frequent commuters or users with consistent riding habits.
  + **Marketing Strategy Implications**
  + Insights from the analysis can inform the design of a new marketing strategy aimed at targeting casual riders and promoting the benefits of annual memberships.
  + Tailored promotional campaigns, targeted messaging, and incentives can be developed to encourage casual riders to make the transition to annual membership, thereby increasing customer retention and long-term profitability for Cyclistic.

**3.1.5 Share Phase**

**Guiding questions**

* Were you able to answer the question of how annual members and casual riders use Cyclistic bikes differently?
* What story does your data tell?
* How do your findings relate to your original question?
* Who is your audience? What is the best way to communicate with them?
* Can data visualization help you share your findings?
* Is your presentation accessible to your audience?

**Key tasks**

* **Determine the Best Way to Share Your Findings**
* Consider the audience and stakeholders who will be receiving the findings.
* Choose the most appropriate format for sharing the insights, such as a written report, presentation slides, or interactive dashboards.
* Ensure that the communication method aligns with the preferences and needs of the audience.
* **Create Effective Data Visualizations**
* Select the most suitable visualization techniques to present the key insights and trends identified in the analysis.
* Use clear and intuitive visualizations, such as bar charts, line plots, and histograms, to effectively convey the information.
* Ensure that the visualizations are easy to interpret and support the narrative of the findings.
* **Present Your Findings**
* Prepare a well-structured presentation or report that highlights the main findings, insights, and recommendations.
* Clearly articulate the methodology used in the analysis and any assumptions made.
* Provide context for the findings and explain the implications for decision-making.
* **Ensure Your Work Is Accessible**
* Make sure that the presentation or report is easily accessible to all stakeholders.
* Consider the accessibility needs of the audience, such as providing alternative formats for individuals with visual impairments.
* Use plain language and avoid technical jargon to ensure that the information is understandable to a wide audience.

**Deliverable**

* **Supporting Visualizations and Key Findings**
* **Average Trip Duration by User Type**
* A bar chart illustrating the average trip duration for annual members and casual riders.
* Key Finding: Annual members tend to have shorter average trip durations compared to casual riders, indicating more frequent and efficient usage patterns.
* **Usage Patterns Over Time**
* A line plot depicting the number of bike trips taken over time, segmented by user type (annual members vs. casual riders).
* Key Finding: Annual members exhibit more consistent usage patterns over time, with higher trip volumes during weekdays, while casual riders show more variability in their riding behavior.
* **Peak Usage Hours**
* A histogram or line plot showing the distribution of bike trip start times throughout the day, separately for annual members and casual riders.
* Key Finding: Annual members tend to use the bike-share service more heavily during peak commuting hours, while casual riders may have more dispersed usage patterns throughout the day.

**3.1.6 Act Phase**

**Guiding questions**

* What is your final conclusion based on your analysis?
* How could your team and business apply your insights?
* What next steps would you or your stakeholders take based on your findings?
* Is there additional data you could use to expand on your findings?

**Deliverable**

* **Top Three Recommendations**
* **Targeted Marketing Campaigns**
* Develop targeted marketing campaigns tailored to the specific needs and preferences of casual riders.
* Highlight the benefits of annual memberships, such as cost savings, convenience, and access to additional features or perks.
* Utilize personalized messaging and incentives to encourage casual riders to upgrade to annual memberships, emphasizing the value proposition for their individual usage patterns.
* **Enhanced User Experience**
* Improve the user experience for both annual members and casual riders by implementing features that address their unique needs.
* For annual members, focus on streamlining the booking process, enhancing bike availability, and providing real-time updates on station status.
* For casual riders, emphasize the flexibility and ease of use of the bike-share program, with clear instructions for one-time rentals and convenient payment options.
* **Community Engagement and Partnerships**
* Foster community engagement and partnerships to increase awareness and adoption of Cyclistic bike-share among potential users.
* Collaborate with local businesses, organizations, and community groups to promote the benefits of cycling and advocate for sustainable transportation options.
* Host events, workshops, or promotional activities to encourage trial usage and showcase the value of annual memberships to target audiences.

These recommendations are based on the analysis of usage patterns and behaviors of Cyclistic bike-share users, aiming to drive conversion of casual riders into annual members while enhancing overall user satisfaction and engagement with the program.

**3.1.7 Visualizations**

A graph of a group of people

Description automatically generated

Fig 3.1.1 Casual vs Members

A graph of different types of bikes

Description automatically generated

Fig 3.1.2 Trips vs Rider Status per Bike Types

A graph of different colored bars

Description automatically generated

Fig 3.1.3 Seasons and Rider Statuses

A graph of different colored bars

Description automatically generated

Fig 3.1.4 Trips per month

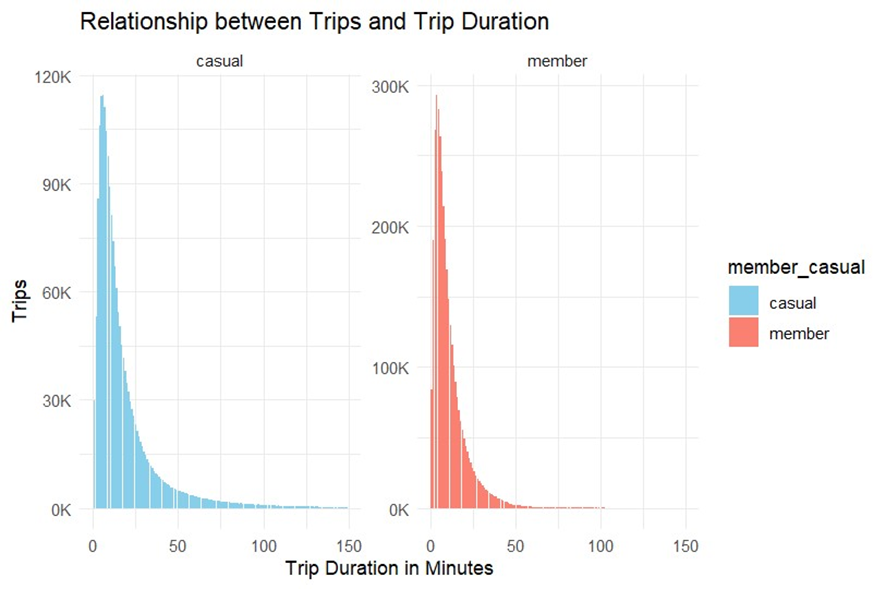


Fig 3.1.5 Trip Duration vs Rider Status

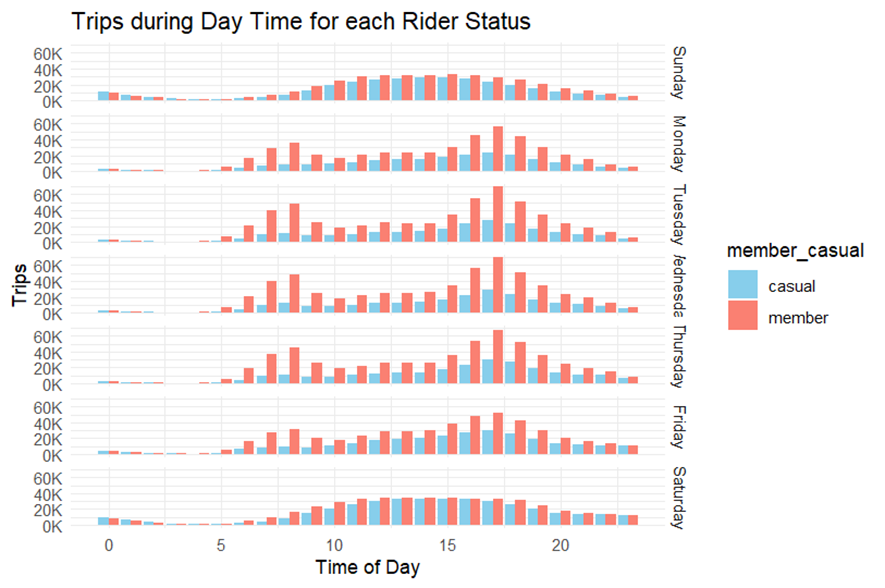


Fig 3.1.6 Trips during daytime

A graph of different colored bars

Description automatically generated

Fig 3.1.7 Average Trip Length on Each Weekday

**4. ACTIVITY LOG**

|  |  |  |  |
| --- | --- | --- | --- |
| **Weeks** | **BrBrief Description of the Weekly Activity** | **Learning Outcome** | **Person In- Charge Signature** |
| Week-1 | Established the foundational knowledge required for data analytics. Learned about the significance of data, big data concepts, and how data analytics drives decision-making processes. | Understanding the fundamentals of data analytics, including terminology and key concepts. Recognition of the importance of data in various industries and contexts. |  |
| Week-2 | Continued to delve into the foundations of data analytics, focusing on asking questions to make data-driven decisions and formulating clear, specific questions and hypotheses. | Understanding the importance of context and audience when framing questions. Practicing asking questions and refining hypotheses based on data insights. |  |
| Week-3 | Data preparation and cleaning techniques, including handling dirty data, performing data cleaning operations, and ensuring data integrity for further analysis. | Exploring tools and software for data preprocessing and cleaning. Gaining practical experience in cleaning real-world datasets to prepare them for analysis. |  |
| Week-4 | Deepen their understanding of data preparation and cleaning, focusing on processing data from dirty to clean, applying various techniques to ensure data quality. | Learning how to structure and format data for analysis. Exploring common data cleaning challenges and how to address them. |  |
| Week-5 | Learning basic data analysis techniques, such as descriptive statistics and exploratory data analysis, to derive insights from data. | Developing skills in summarizing and visualizing data to extract meaningful insights. Learning how to formulate and answer analytical questions using data. |  |
| Week-6 | Explore data analysis techniques and visualization methods, focusing on sharing data through the art of visualization and creating clear and compelling visualizations to communicate insights effectively. | Learning principles of effective data visualization and storytelling. Exploring techniques for creating clear and compelling data visualizations. |  |
| Week-7 | Applying programming languages like R and Python for data analysis, focusing on manipulating data, conducting statistical analysis, and creating visualizations using these tools. | Learning how to manipulate and analyze data using R and Python libraries. Exploring statistical analysis techniques and applying them to real datasets. |  |
| Week-8 | Deepen their understanding of data analysis using R and Python, focusing on data analysis with Python, including data manipulation, statistical analysis, and visualization using Python libraries. | Learning advanced data manipulation and analysis techniques using Python and pandas library. Exploring data visualization libraries such as Matplotlib and Seaborn. |  |
| Week-9 | Exploring advanced data analytics techniques, including data visualization with Tableau. The learner learned how to design interactive dashboards and create compelling visualizations to explore and present data effectively. | Mastering advanced data visualization techniques using Tableau software. Understanding how to design interactive dashboards to explore and present data effectively. |  |
| Week-10 | Worked on completing a capstone project. Applied all the skills and knowledge acquired throughout the program to solve a real-world data analytics challenge, focusing on data analysis, visualization, and interpretation. | Applied data analytics techniques, visualization skills, and storytelling principles to solve a real-world data analytics problem. Demonstrated proficiency in data analysis, interpretation, and presentation of findings. |  |
| Week-11 | Project Review | Improvements in the implementation and better communication in storytelling. |  |
| Week-12 | Feedback on the internship | Provided valuable feedback on the internship. |  |

Table 4.1 Weekly Activity Log

**5. CONCLUSION**

In recent years, the field of data analytics has witnessed significant technological advancements that have reshaped the way data analysts operate. Firstly, the proliferation of big data technologies like Hadoop and Apache Spark has enabled analysts to handle vast volumes of data efficiently. These distributed computing frameworks allow for parallel processing across distributed systems, making it feasible to tackle complex analytics tasks at scale. Additionally, cloud computing platforms such as AWS, Azure, and GCP have democratized access to computing resources, providing scalable storage, and computing capabilities on-demand, thus streamlining data analysis workflows, and reducing infrastructure costs.

Furthermore, the integration of machine learning and artificial intelligence technologies into data analytics tools has revolutionized the way insights are extracted from data. Machine learning algorithms can automatically detect patterns, trends, and anomalies in large datasets, enhancing the efficiency of analysis. Advanced data visualization tools like Tableau and Power BI have also empowered data analysts to create interactive and visually compelling dashboards and reports, enabling them to communicate insights effectively to stakeholders. These developments highlight the evolving landscape of data analytics, where emerging technologies continue to drive innovation and enable data analysts to derive actionable insights from increasingly complex datasets

**REFERENCES**

1. <https://www.coursera.org/articles/data-analytics>
2. <https://www.w3schools.com/r/r_intro.asp>
3. <https://www.python.org/>
4. <https://www.coursera.org/professional-certificates/google-data-analytics>
5. <https://www.geeksforgeeks.org/sql-tutorial/>
6. <https://pandas.pydata.org/>
7. <https://numpy.org/>
8. <https://matplotlib.org/>
9. <https://seaborn.pydata.org/>