“Harnessing data visualization to generate strategic insights at Fipchain Technologies Pvt Ltd”

Internship report submitted to Sastra Deemed University

in partial fulfillment of the requirements for the degree of

MASTER OF BUSINESS ADMINISTRATION

Submitted by

PRIYADHARSHINI M

(Reg.No:125071061)

Under the guidance of

DR. P. SUGANTHI, Ph.d.,

Sr. Asst. Professor, School of Management

SASTRA Deemed University, Thanjavur -613401

SCHOOL OF MANAGEMENT

SASTRA DEEMED UNIVERSITY

THANJAVUR-613401

JUNE-2024

OFFER LETTER

fipchain

FIPCHAIN Technology Pvt Ltd
formerly known as NFT Creator Economy Private Limited
CIN: U72200TN2021PTC146604

Internship Offer Letter

Role: MBA Intern – Analyst (Data Analyst and BI)

Hi Priyadarshini,

In this role, you will be responsible for managing and optimizing the frontend of our Power BI platform. You will work closely with stakeholders to understand their reporting needs, design visually appealing and intuitive dashboards, and ensure the platform meets user requirements. This is an excellent opportunity for a driven individual to gain hands-on experience in data visualization and business intelligence while making a tangible impact on our organization.

Key Responsibilities:

- Collaborate with stakeholders to gather requirements and understand business needs for reporting and data visualization.
- Design and develop interactive and visually appealing dashboards using Power BI to present complex data in an understandable manner.
- Customize and optimize user interfaces to enhance user experience and ensure ease of navigation.
- Work closely with the backend team to integrate frontend components with backend datasets and ensure data accuracy and consistency.
- Conduct user training sessions and provide support to users on navigating and interpreting data in the Power BI platform.
- Stay updated on industry best practices and emerging trends in data visualization and business intelligence to continuously improve the frontend of the Power BI platform.
- Collaborate with cross-functional teams to troubleshoot and resolve any frontend issues or performance bottlenecks.

Renumeration:

You will be paid Rs.5,000 as monthly stipend for this role and the continuity of the internship is solely dependent on the work update you put forth from time to time and this letter does not guarantee any fixed term.

Regards,

GK Tirunavukarasu
Founder & CEO

CERTIFICATE

This is to certify that this internship report titled **"Harnessing data visualization to generate strategic insights at Fipchain Technologies Pvt Ltd"** is a Bonafide work done by **PRIYADHARSHINI M (Reg. No: 125071061)** for the academic year 2023-2024 in partial fulfilmentof the requirements for the award of the Degree of Master of Business Administration.

Signature of Guide

Dr. P. Suganthi,Ph.d
Sr. Asst. Professor
SASTRA Deemed University
Thanjavur-613401

Signature of Dean

DR. S. Selvabaskar, Ph.d
Dean, School of Management
SASTRA Deemed University
Thanjavur-613401

Submitted for Internship Viva-Voce examination held on _____

Examiner-I

Examiner-II

DECLARATION

I hereby declare that this internship report titled titled **“Harnessing data visualization to generate strategic insights at Fipchain Technologies Pvt Ltd”** submitted by me for the academic year 2023-2024 in partial fulfilment of the requirements for the award of the Degreeof Master of Business Administration. This is my original work and the project has not formedthe basis for the award of any degree, diploma, or other similar titles.

Place: Thanjavur

Date:

Student Signature

(**PRIYADHARSHINI M)**

ACKNOWLEDGEMENTS

First, I would like to give my immense gratitude to the almighty God for his blessing and guidance bestowed on me throughout this research study.

I take this opportunity to acknowledge my deep sense of gratitude to our **Honourable Vice-Chancellor, Dr.S.Vaidhyasubramaniam,** SASTRA Deemed University for having allowed me to pursue this MBA programme.

I would like to express my sincere thanks to **Dr. S.Selvabaskar, Dean, School of Management,** SASTRA Deemed University for his valuable suggestions and support.

I would like to express my heartful thanks to **Dr.Suganthi P, Asst. Professor, School of Management,** SASTRA Deemed University for giving valuable input and concrete encouragement to complete this work successfully.

I would like to express my sincere gratitude to the team at **Fipchain Technology Pvt Ltd** for providing me with this enriching internship experience. A special note of thanks goes to Mr. **G.K. Thirunavukarasu** for granting me this wonderful opportunity.

I am particularly grateful to **Loganathan** Sir, my mentor, whose insights and encouragement have been invaluable in shaping my learning journey.

I would also like to extend my heartfelt thanks to the CEO Club cohort and the startup community associated with SASTRA Deemed University. Their innovative ideas, collaborative spirit, and dynamic approach have greatly inspired me and enriched my overall experience.

TABLE OF CONTENTS

EXECUTIVE SUMMARY

This internship experience as a Frontend Data Analyst at Fipchain Technology Pvt Ltd from May to June 2024. During this period, the primary focus was on utilizing Power BI to conduct detailed analysis and visualization of a diverse movie dataset sourced through extensive web scraping efforts. The project involved cleaning and preparing the dataset to ensure data integrity before embarking on an in-depth Exploratory Data Analysis (EDA). This process was aimed at uncovering meaningful insights into various aspects of the film industry. This included exploring movie genres, evaluating ratings based on storyline and performances, examining production house affiliations, and identifying the key factors influencing both box office success and critical acclaim. By meticulously scrutinizing the relationships between different variables within the dataset, able to identify trends and patterns that could significantly inform strategic decision-making processes within the industry. One of the key responsibilities undertook was the cleaning and preparation of the data, ensuring that the dataset was free from inconsistencies and ready for detailed analysis. Following this, employed Power BI to perform statistical analyses, creating visual representations that highlighted important findings. This process allowed to gain a deeper understanding of the complexities involved in the entertainment sector, particularly how data-driven insights can play a pivotal role in shaping business strategies. Through this hands-on experience, developed a strong proficiency in data visualization and analytical reasoning, essential skills that are critical for future growth in the field of data analytics. The internship not only enhanced the technical capabilities but also provided with practical insights into the strategic implications of data analytics, particularly how they can drive organizational success in a dynamic industry landscape.The internship at Fipchain Technology Pvt Ltd was a transformative journey. It not only enriched the technical knowledge but also deepened the understanding of how data analytics can be leveraged to derive actionable insights, ultimately facilitating informed decision-making processes. This experience has been instrumental in shaping the career aspirations, equipping with the tools and knowledge necessary to excel in the field of data analytics, particularly within industries that are as complex and multifaceted as Film and Entertainment.

INTRODUCTION

The core focus of the experience at Fipchain Technology Pvt Ltd was to gain practical insights into analyzing large datasets and leveraging these insights to drive strategic business decisions. Throughout this internship, deepened the understanding of Exploratory Data Analysis (EDA) and honed the skills in applying statistical methods using tools like Jupyter Notebooks, Power BI, and Excel, significantly enhancing the proficiency in data analytics.

One of the key projects involved analyzing a detailed movie dataset, which provided hands-on experience in cleaning and preparing data. The dataset included information such as movie titles, release dates, genres, ratings, production house affiliations, and box office figures. They employed various data cleaning techniques to address inconsistencies, missing values, and duplications, ensuring the data was ready for in-depth examination. This experience played a crucial role in converting complex datasets into accessible and meaningful information to support the organization's strategic decisions.

Initially, they were responsible for evaluating customer relationship management (CRM) tools such as Zoho CRM, HubSpot, and Microsoft Dynamics. The team analyzed costs, evaluated service varieties, and compared these tools based on several criteria, including cost, range of services, customer support, data privacy, and quality of service. This comprehensive analysis enabled them to identify the most suitable CRM tool for the company, enhancing operational efficiency and supporting growth objectives.

In addition to CRM tool evaluation, they focused on performing Exploratory Data Analysis (EDA) to uncover underlying patterns, trends, and relationships within the data. This included examining variables such as the release year, genre, ratings, production house, box office success, and IMDb ratings. This analytical approach provided practical insights into the dynamics of the film industry and enhanced their proficiency in data cleaning, statistical analysis, and data visualization. Data entry and research were also integral components of their comprehensive data analytics experience during this internship.

DATA VISUALIZATION

Data visualization in data analytics is the process of representing data and information in graphical formats, such as charts, graphs, maps, and dashboards, to make complex data more understandable, accessible, and actionable. It enables the identification of patterns, trends, correlations, and outliers within data, thereby facilitating more informed decision-making and effective communication of insights to both technical and non-technical stakeholders.

The importance of data visualization in data analytics cannot be overstated, as it plays a pivotal role in making data understandable, actionable, and accessible to a wide range of audiences. Here's a detailed explanation of its significance:

1. Simplifies Complex Data:

Data often comes in large, complex sets that can be difficult to interpret when presented in raw formats such as tables or spreadsheets. Data visualization transforms this raw data into visual formats like charts, graphs, and maps, which makes it easier to digest.

2. Enhances Insight Discovery:

Data visualization is a powerful tool for discovering insights within data. By representing data visually, patterns, trends, and correlations that might not be immediately obvious in raw data become much more apparent. For instance, a scatter plot might reveal a strong correlation between two variables, or a heat map might show geographic regions with higher activity levels. This ability to uncover hidden insights is crucial for making informed decisions and driving strategy.

3. Improves Decision-Making:

Effective decision-making relies on understanding data accurately and quickly. Data visualizations provide a clear and concise summary of the data, enabling decision-makers to assess situations, identify opportunities or risks, and make informed choices swiftly.

4. Facilitates Communication:

One of the most critical aspects of data visualization is its ability to communicate data-driven insights to a broader audience, including those who may not have technical expertise. Visuals are often more engaging and easier to understand than text or numbers alone. For instance,

presenting a complex data analysis in the form of a well-designed infographic can convey the main points effectively to stakeholders, ensuring that the message is clear and understood.

5. Increases Engagement:

Interactive data visualizations allow users to engage with the data in a more meaningful way. Tools like Tableau, Power BI, and Google Data Studio enable users to explore data by filtering, drilling down, and interacting with visual elements. This interactivity not only increases user engagement but also empowers users to explore the data on their own, leading to a deeper understanding and more personalized insights.

6. Identifies Outliers and Anomalies:

Outliers and anomalies in data can be critical indicators of underlying issues or opportunities. Data visualization makes it easier to spot these irregularities. For example, a box plot can quickly highlight data points that deviate significantly from the norm, signaling potential errors, fraud, or exceptional cases that require further investigation. Identifying these outliers can lead to more accurate analyses and better-informed decisions.

7. Supports Data-Driven Culture:

In organizations striving to build a data-driven culture, data visualization plays a key role. It makes data accessible to everyone, not just data scientists or analysts, fostering an environment where decisions are based on data rather than intuition or guesswork. Regularly sharing visual dashboards and reports can help inculcate this culture, encouraging teams to use data as the basis for their daily work and strategic planning.

8. Aids in Storytelling:

Data visualization is also a powerful tool for storytelling. It allows data analysts and business leaders to craft compelling narratives around the data, turning abstract numbers into stories that resonate with the audience. A well-crafted story supported by visuals can not only inform but also persuade and inspire action. For instance, a series of charts and graphs can be used to tell the story of a company's growth, challenges, and future opportunities in a way that is both informative and engaging.

9. Enhances Data Accessibility:

Not all stakeholders have the time or expertise to analyze raw data. Data visualization bridges this gap by presenting data in a format that is accessible to everyone, regardless of their technical background. This democratization of data means that insights are not confined to data specialists but are available to all decision-makers, enabling a more collaborative and informed approach to business strategy.

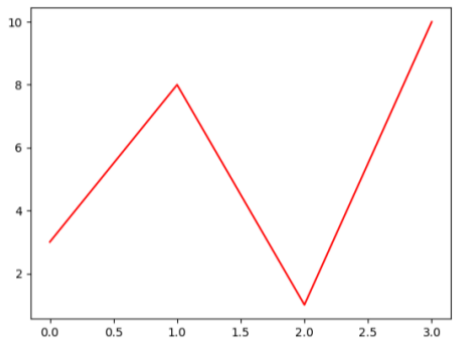
10. Supports Real-Time Monitoring:

In environments where real-time data is critical, such as finance, healthcare, or supply chain management, data visualization tools allow for the real-time monitoring of key metrics. Dashboards that update in real-time provide an immediate overview of the current situation, helping organizations respond quickly to changes, capitalize on emerging opportunities, or mitigate risks.

Common types of Data Visualizations

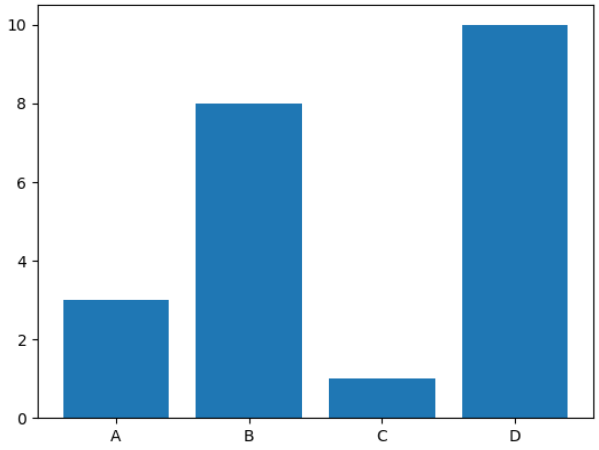
1. Line Plot

- A line plot is used to show trends over time or continuous data.
- A line plot connects individual data points with a continuous line, making it easy to see how a particular variable changes over a period.
- The x-axis typically represents time (such as years, months, days, etc.), while the y-axis represents the variable being measured (like sales, stock prices, etc.).
- Line plots are effective for visualizing time series data and spotting trends, cycles, or fluctuations in the data.



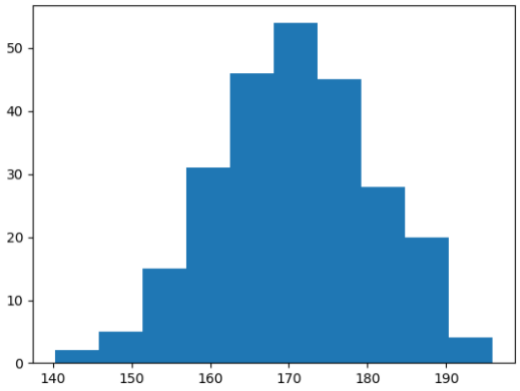
2. Bar Chart

- A bar chart is used to compare quantities across different categories.
- It consists of rectangular bars, where the length of each bar is proportional to the value it represents.
- The categories are typically placed on the x-axis, while the values are on the y-axis (or vice versa).
- Bar charts can be vertical or horizontal and are ideal for comparing different groups or categories side by side.



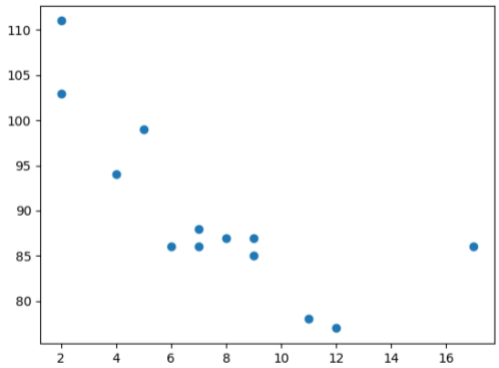
3. Histogram

- A histogram displays the distribution of a dataset by grouping data into bins.
- The x-axis represents the range of values (bins), while the y-axis shows the frequency (the number of data points within each bin).
- Histograms help to visualize the shape of the data distribution, such as whether it's normal (bell-shaped), skewed, bimodal, etc.
- Unlike bar charts, which compare different categories, histograms are used for continuous data and show how the data is distributed across different intervals.



4. Scatter Plot

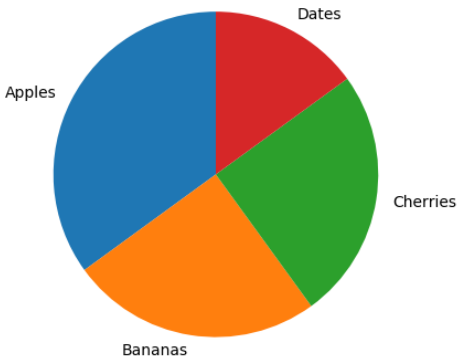
- A scatter plot shows the relationship between two variables by displaying points at the intersection of their values.
- Each point on the scatter plot represents an observation, with one variable plotted on the x-axis and the other on the y-axis.
- Scatter plots are useful for identifying correlations or relationships between variables, such as linear, exponential, or no relationship at all.
- They can also reveal patterns, clusters, and outliers within the data.



5. Pie Chart

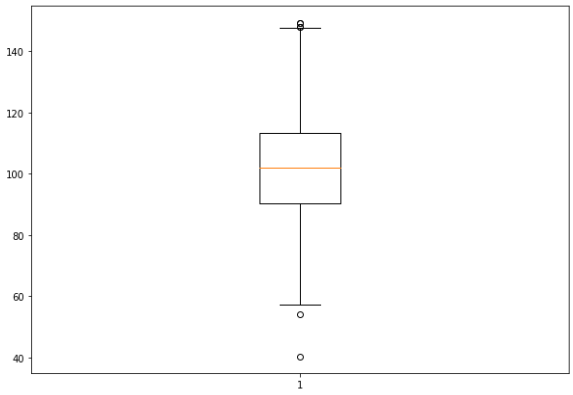
- A pie chart shows the proportions of a whole by dividing a circle into slices representing different categories.
- Each slice of the pie represents a category, with the size of the slice proportional to the percentage or proportion that category contributes to the total.

- Pie charts are ideal for showing the composition of a dataset and comparing the relative sizes of the parts to the whole.
- They are most effective when there are only a few categories, as too many slices can make the chart hard to interpret.



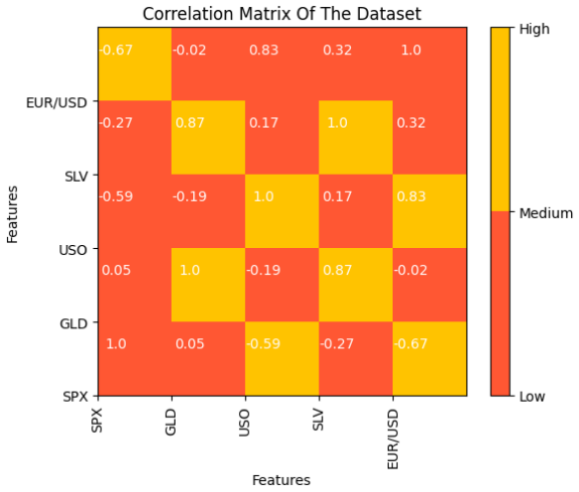
6. Box Plot

- A box plot (or box-and-whisker plot) summarizes the distribution of a dataset by showing the median, quartiles, and potential outliers.
- The box represents the interquartile range (IQR), which is the middle 50% of the data, with the line inside the box representing the median (the middle value).
- The "whiskers" extend from the box to show the range of the data, excluding outliers.
- Outliers, or data points that are significantly higher or lower than the rest of the data, are often plotted as individual points outside the whiskers.
- Box plots are useful for comparing distributions between different groups or datasets.



7. Heatmap

- A heatmap represents data in a matrix format, where individual values are displayed as colors.
- The matrix format of a heatmap typically has categories on both the x-axis and y-axis, and the cells within the matrix are colored according to the values they represent.
- The color gradient is used to show the magnitude of the values, making it easy to spot patterns, correlations, or areas of interest.
- Heatmaps are particularly useful for visualizing large datasets and for spotting trends, variations, or anomalies within the data.



DATA CLEANING

Data cleaning in data analytics is the process of identifying, correcting, or removing errors, inconsistencies, and inaccuracies in a dataset to ensure that the data is accurate, complete, and ready for analysis. This involves handling missing data, correcting erroneous entries, resolving duplicates, and standardizing formats, all of which are essential steps to improve the quality of the data and ensure reliable and valid analytical results.

The importance of data cleaning in data analytics is crucial because it directly impacts the quality, accuracy, and reliability of the analysis. Here's why data cleaning is essential:

1. Ensures Data Accuracy:

Data cleaning corrects errors and inaccuracies in the dataset, such as typos, incorrect entries, and inconsistencies. Accurate data is fundamental to producing valid results in any analysis. If the data is flawed, the insights derived from it will be misleading or incorrect, leading to poor decision-making.

2. Improves Data Consistency:

Inconsistent data, such as varying formats for dates or differing spellings of the same entity, can cause confusion and errors in analysis. Data cleaning standardizes these inconsistencies, ensuring that all data is in a uniform format, which simplifies analysis and improves comparability across different datasets.

3. Eliminates Redundancies:

Duplicate records and redundant data entries can distort analysis by giving undue weight to certain data points. Data cleaning identifies and removes these duplicates, ensuring that each data point is counted only once, leading to more accurate analysis.

4. Handles Missing Data:

Missing data can skew results if not addressed properly. Data cleaning involves strategies to handle missing data, such as imputing values, filling gaps, or excluding incomplete records, depending on the context. Proper handling of missing data ensures that the analysis is as complete and accurate as possible.

5. Enhances Data Quality:

High-quality data is essential for trustworthy analytics. Data cleaning improves the overall quality of the data by ensuring it is complete, accurate, and free from errors. High-quality data leads to more reliable insights and better decision-making.

6. Boosts Efficiency in Analysis:

Clean data simplifies the analysis process, as analysts can focus on extracting insights rather than spending time correcting errors or dealing with inconsistencies. This efficiency leads to faster and more accurate analysis, saving time and resources.

7. Enables Better Decision-Making:

Reliable data is the foundation of sound decision-making. Clean data ensures that decisions are based on accurate, consistent, and complete information, reducing the risk of errors and improving the outcomes of business strategies and operations.

8. Facilitates Data Integration:

When combining data from multiple sources, data cleaning is essential to ensure that the data from different systems can be integrated seamlessly. Clean data makes it easier to merge datasets and ensures that the combined data is consistent and accurate.

9. Increases Confidence in Results:

Clean, accurate data increases the confidence that stakeholders have in the results of the analysis. When data is clean, the insights derived are more credible, which is crucial for gaining trust from decision-makers and other stakeholders.

EXPLORATORY DATA ANALYSIS

Exploratory Data Analysis (EDA) in data analytics is the process of examining and visualizing a dataset to summarize its main characteristics, often using statistical graphics, plots, and information tables. EDA is a critical early step in data analysis that helps analysts understand the data's structure, identify patterns, spot anomalies, test hypotheses, and check assumptions before moving on to more complex modeling or predictive analytics.

1. Understanding the Data

- EDA provides a comprehensive overview of the dataset, helping analysts understand its structure, distribution, and characteristics.
- **Importance:** Without a thorough understanding of the data, any analysis or modeling built on it can be flawed. EDA ensures that analysts are familiar with the data types, range of values, and the relationships between variables, which is essential for making informed decisions later in the process.

2. Identifying Patterns and Relationships

- EDA helps in discovering underlying patterns, trends, and correlations within the data that may not be immediately apparent.
- **Importance:** These insights can reveal important relationships between variables that can guide the selection of features for predictive modeling or other analyses. Understanding these patterns is crucial for developing accurate models and making reliable predictions.

3. Detecting Anomalies and Outliers

- EDA is used to identify outliers and anomalies that could skew results or indicate errors in the data.
- **Importance:** Detecting and understanding outliers is vital because they can significantly impact the results of an analysis. Anomalies might indicate data entry errors, fraud, or unique cases that require special attention. Handling these appropriately during EDA ensures that they do not adversely affect the outcome of the analysis.

4. Assessing Data Quality

- EDA involves examining the data for issues such as missing values, inconsistencies, and inaccuracies.
- **Importance:** Data quality directly impacts the validity of the analysis. EDA helps in identifying data quality issues early on, allowing analysts to clean and preprocess the data before applying more advanced techniques. This ensures that the analysis is based on accurate and reliable data, leading to more trustworthy results.

5. Informing Feature Selection and Engineering

- EDA helps in selecting the most relevant variables (features) for analysis and identifying opportunities for creating new features.
- **Importance:** The choice of features greatly influences the performance of predictive models. EDA guides feature selection by highlighting which variables are most strongly correlated with the outcome of interest. Additionally, EDA may reveal opportunities for

feature engineering, such as creating interaction terms or combining variables in a meaningful way.

6. Guiding Further Analysis

- The insights gained from EDA help shape the direction of further analysis, including hypothesis testing, model selection, and evaluation.
- **Importance:** EDA is not just a preliminary step; it informs every subsequent stage of the data analysis process. By revealing the data's underlying structure and relationships, EDA helps analysts choose the appropriate analytical methods and models, ensuring that the analysis is well-targeted and effective.

7. Reducing Bias and Misleading Results

- EDA helps in detecting biases or misrepresentations in the data that could lead to incorrect conclusions.
- **Importance:** Data that has biases or is misrepresentative can produce misleading results. EDA allows analysts to identify and address these issues before they influence the analysis, thereby reducing the risk of drawing incorrect conclusions and making poor decisions based on faulty data.

8. Enhancing Communication and Reporting

- EDA involves creating visualizations and summaries that communicate the data's key characteristics.
- **Importance:** Effective communication of data insights is crucial, especially when presenting findings to stakeholders who may not have a deep technical background. EDA provides the tools for creating clear, compelling visualizations that can convey complex data insights in an accessible way, making it easier to communicate findings and drive data-driven decision-making.

Common Tools and Techniques in EDA

- **Descriptive Statistics**: Summarizing data using measures like mean, median, mode, standard deviation, and variance.

- **Data Visualization**: Using plots such as histograms, box plots, scatter plots, and pair plots to visually explore the data.
- **Correlation Analysis**: Checking the relationships between variables using correlation coefficients.
- **Data Cleaning**: Identifying and handling missing values, duplicates, and errors in the data.
- **Feature Engineering**: Creating new variables or modifying existing ones to improve the performance of predictive models.

STATISTICAL ANALYSIS

Statistical analysis in data analytics refers to the process of collecting, exploring, and interpreting data to identify patterns, trends, relationships, and underlying structures. It involves the application of statistical methods to understand and quantify data characteristics, make inferences, and support decision-making based on empirical evidence. Statistical analysis is fundamental in data analytics, as it provides the mathematical foundation for understanding data and deriving meaningful insights.

Key Components of Statistical Analysis

1. **Descriptive Statistics**
    - Descriptive statistics summarize and describe the main features of a dataset through measures such as mean, median, mode, standard deviation, variance, and percentiles.
2. **Inferential Statistics**
    - Inferential statistics use a sample of data to make inferences or predictions about a larger population. This includes hypothesis testing, confidence intervals, and regression analysis.

3. **Hypothesis Testing**

    - Hypothesis testing is a method for testing a claim or hypothesis about a parameter in a population, using sample data.

4. **Correlation and Regression Analysis**

    - Correlation analysis measures the strength and direction of the relationship between two variables, while regression analysis models the relationship between dependent and independent variables.

5. **Probability Distribution**

    - Probability distributions describe how the values of a random variable are distributed. Common distributions include normal, binomial, and Poisson distributions.

6. **ANOVA (Analysis of Variance)**

    - ANOVA is used to compare the means of three or more groups to determine if there are statistically significant differences between them.

7. **Time Series Analysis**

    - Time series analysis involves analyzing data points collected or recorded at specific time intervals to identify trends, seasonal patterns, and cyclical movements.

Importance of Statistical Analysis

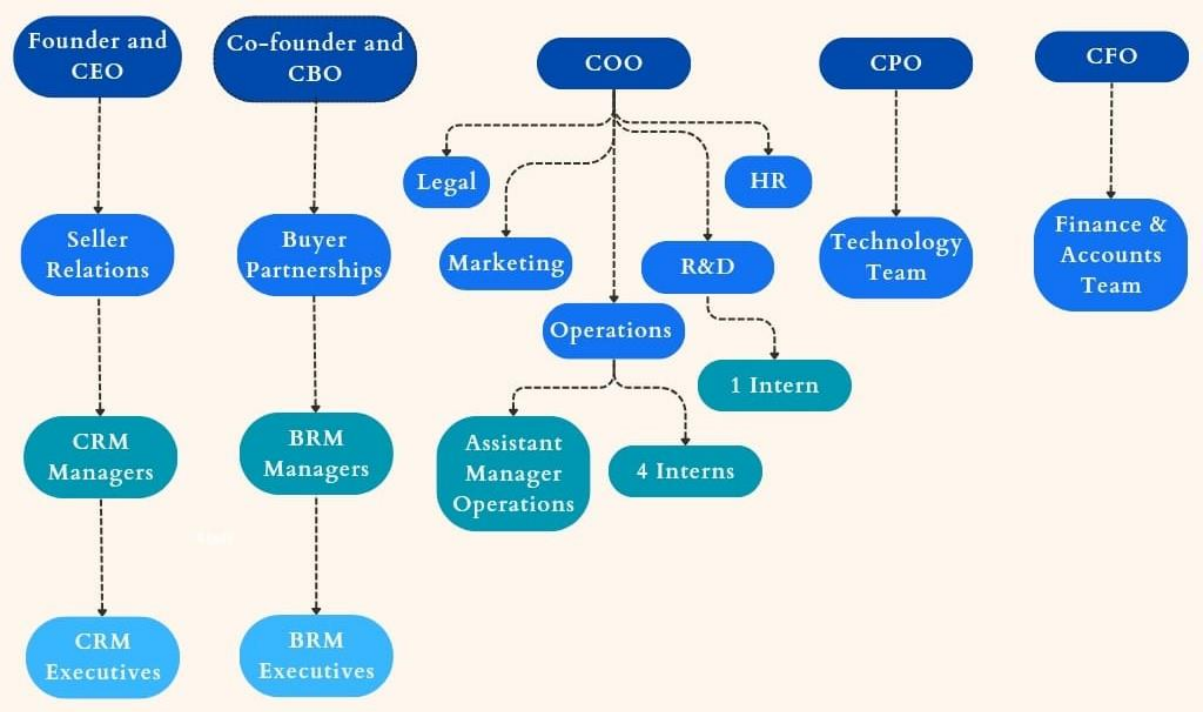
- **Data Understanding**: Statistical analysis helps in understanding the underlying structure of the data, including central tendencies, variability, and distribution.
- **Pattern Identification**: By analyzing data statistically, patterns, correlations, and trends can be identified, which may not be immediately visible.
- **Validation of Findings**: Statistical analysis allows for the testing of hypotheses and validation of findings, ensuring that conclusions drawn from data are reliable.

COMPANY OVERVIEW

Fipchain Technology Pvt Ltd has emerged to revolutionize the unexplored facets of the Film and Entertainment industry by offering innovative B2B, B2C, and Fintech solutions. One of its key platforms, Producerbazaar.com, is a cutting-edge marketplace that merges WEB2 and WEB3 technologies. It seamlessly integrates blockchain to connect creators, producers, aggregators, and buyers for the purchase and sale of movie-related Intellectual Property rights globally.

Website : https://www.producerbazaar.com/
Industry : Technology, Information and Media
Founded : 2021

ORGANISATION STRUCTURE



POSITION

Frontend data analyst

During the internship at Fipchain Technology Pvt Ltd, they held the role of Frontend Data Analyst, focusing primarily on utilizing Jupyter Notebooks and Power BI to analyze and visualize data related to movie datasets.

DESCRIPTION OF DEPARTMENT/BRANCH ACTIVITIES/OPERATIONS

They were part of the Power BI team, focusing on using data visualization tools to derive actionable insights from complex datasets. This role was crucial in transforming raw data into accessible and meaningful information for strategic decision-making within the organization.

RESPONSIBILITIES ASSIGNED

Initially, they were part of the frontend team, where they learned to use various CRM tools such as Zoho CRM, HubSpot, and Microsoft Dynamics. Their tasks included acquiring and analyzing the costs quoted for these services and evaluating the variety of services offered. They compared these CRM tools based on several criteria, including cost, range of services, customer support, data privacy, credibility, and quality of service.

Additionally, they assessed the integration capabilities of these CRM tools with other software used by the company, and evaluated their user-friendliness and scalability. They also reviewed how each CRM tool's features aligned with the company's specific business needs and goals. This comprehensive analysis enabled them to identify the most suitable CRM tool, ensuring the selection of a solution that would enhance operational efficiency and support growth objectives. This experience broadened their technical knowledge and improved their analytical and decision-making skills in a real-world business context.

Their primary responsibility was to clean and analyze a comprehensive movie dataset to assess relationships between variables, ensuring data integrity and preparing it for detailed analytical examination.

DETAILS OF TASKS COMPLETED

Data cleaning and preparation

The initial phase of the project focused on data cleaning and preparation, a critical step to ensure data integrity. The movie dataset contained extensive information, including movie titles, release dates, genres, ratings for storyline and performances, production house affiliations, and box office figures. However, the raw data had inconsistencies, missing values, and duplications.

They employed various data cleaning techniques to address these issues. This included:

1. **Handling Missing Values:** Identifying and filling missing data points using appropriate methods such as mean imputation for numerical data and mode imputation for categorical data.
2. **Data Standardization:** Ensuring uniformity in data entries, especially in categorical fields like genres and production houses, to facilitate accurate analysis.
3. **Outlier Detection and Treatment:** Identifying outliers that could skew the analysis and deciding whether to retain or remove them based on their relevance and impact on the overall dataset.
4. **Data Transformation:** Converting data into suitable formats for analysis, such as parsing dates and normalizing text entries.

This meticulous data cleaning process was pivotal in preparing the dataset for in-depth analysis, ensuring that the insights derived would be reliable and actionable.

Data entry and Research

Data entry and research were also key components of their internship. While web scraping provided the bulk of the dataset, they supplemented this data with additional information sourced from various industry reports, databases, and credible online resources. This involved:

1. **Web Scraping:** Extracting large volumes of data from movie databases and review sites using Python scripts and tools like BeautifulSoup and Scrapy.

2. **Manual Data Entry:** Incorporating data that could not be scraped automatically, ensuring accuracy and consistency.
3. **Secondary Research:** Conducting extensive research to gather contextual information about the movies, such as production budgets, marketing expenditures, and audience demographics.

This comprehensive approach to data collection enriched the dataset, allowing for a more nuanced and complete analysis of the movie industry.

Exploratory Data Analysis (EDA)

With the dataset cleaned and prepared, they performed Exploratory Data Analysis (EDA) using Power BI. EDA is a crucial step in the data analysis process, helping to uncover underlying patterns, trends, and relationships within the data. One of the significant projects they undertook was conducting EDA on a movie dataset compiled through meticulous web scraping.The dataset encompassed various critical variables:

- **Web Scraper Order:** Sequential order of movie data acquisition.
- **Year of Release:** Release year of the movies.
- **Title:** Movie title.
- **Primary Genre:** Main genre of the movie (e.g., Action, Drama, Comedy).
- **Story Rating, Script/Screenplay Rating, Hero Rating, Heroine Rating, Director Rating, Music Director Rating:** Ratings for different aspects of the movie.
- **Production House:** Production company responsible for the movie.
- **Writer:** Writer(s) credited for the movie.
- **Success Due to Star Value:** Indicator of whether the movie's success was influenced by its cast's star power.
- **Box Office Success:** Financial success of the movie.
- **Recognition:** Awards or critical acclaim received by the movie.
- **IMDB Rating:** IMDb rating of the movie.

The EDA process involved scrutinizing these variables to uncover trends and patterns that could illuminate factors contributing to a movie's success or critical reception. This analytical approach not only enhanced my proficiency in data cleaning and statistical analysis but also provided practical insights into the dynamics of the film industry.

DATE	TASK / ACTIVITY
24.05.24	Attended a meeting on understanding the CRM tool, where key features and functionalities were discussed.
25.05.24	Initiated the learning process for various CRM tools, with a particular emphasis on Microsoft 365.
27.05.24	Gained a comprehensive understanding of the structural framework of Microsoft 365
28.05.24	Conducted a comparative analysis of Microsoft 365 with other CRM tools, such as Zoho, to identify differences.
29.05.24	Continued in-depth exploration of the workings of Microsoft 365, focusing on its integration and application.
01.06.24	Delivered a detailed presentation on CRM tools, highlighting key insights and comparative findings.
03.06.24	Received the assignment to analyze a movie dataset.
04.06.24	Conducted an initial analysis of the movie dataset, identifying key variables and their potential relationships.
05.06.24	Engaged in an exploratory analysis of the dataset to familiarize with the data and its attributes.
06.06.24 - 07.06.24	Completed the data cleaning process, ensuring the dataset.
08.06.24	Discuss the current status of the data analysis project in the meeting.
10.06.24	Began detailed descriptive analysis of each column within the dataset, interpreting their significance.
11.06.24 - 13.06.24	Utilized data visualization techniques in jupyter notebook,to identify relationships between box office success and other variables.
14.06.24	Performed statistical analysis using the Chi-square test to examine the relationship between box office success and other key variables.
15.06.24	Attended a progress meeting to report findings and outline the next steps in the analysis process.
17.06.24 - 19.06.24	Conducted data visualization in Jupyter Notebook to explore the relationship between IMDb ratings and other variables.

20.06.24	Executed statistical analysis using the Chi-square test to assess the relationship between IMDb ratings and other variables.
21.06.24	Prepared a PowerPoint presentation to summarize and communicate the findings from the analysis.
22.06.24	Attended a progress meeting to review the presentation content.
24.06.24 - 26.06.24	Applied data visualization techniques in Jupyter Notebook to explore the relationship between recognition and other variables.
27.06.24	Conducted statistical analysis using the Chi-square test to evaluate the relationship between recognition and other variables.
28.06.24	Finalized and polished the presentation, ensuring all key findings were clearly communicated.
29.06.24	Presented the comprehensive analysis, detailing the insights derived from the data and the implications of the findings.

LEARNING OUTCOMES

CRM tools

1. Comprehensive Understanding of CRM Tools:

- Gained in-depth knowledge of various CRM tools such as Zoho CRM, HubSpot, and Microsoft Dynamics.
- Understood the detailed features and functionalities of each CRM tool, including their strengths and limitations.

2. Cost-Benefit Analysis Skills:

- Developed the ability to compare costs and benefits of different CRM tools, ensuring the selection of the most cost-effective option without compromising quality.

3. Service Range Evaluation:

- Learned to evaluate the range of services offered by different CRM tools, ensuring that the chosen tool meets specific business needs.

4. Customer Support Assessment:

- Gained experience in assessing the quality of customer support provided by various CRM tools, ensuring reliable and efficient assistance when needed.

5. Data Privacy and Security Analysis:

- Understood the importance of data privacy and security in CRM tools and learned to evaluate these aspects to ensure the protection of sensitive company data.

6. Credibility and Quality of Service:

- Developed skills to assess the credibility and overall quality of service provided by CRM tools, ensuring a seamless and efficient user experience.

7. Integration Capabilities:

- Learned to evaluate the integration capabilities of CRM tools with other software used by the company, ensuring a cohesive and efficient workflow.

8. User-Friendliness and Scalability:

- Gained the ability to assess the user-friendliness and scalability of CRM tools, ensuring they can grow with the company's needs and remain easy to use.

9. Alignment with Business Goals:

- Learned to align CRM tool features with specific business goals and objectives, enhancing operational efficiency and supporting growth.

10. Enhanced Analytical and Decision-Making Skills:

- Improved analytical and decision-making skills through the comprehensive analysis of CRM tools, ensuring informed and strategic choices.

Data entry

1. Data Accuracy and Consistency:

- Developed the ability to ensure data accuracy by meticulously entering data without errors.
- Learned to maintain consistency in data entries, ensuring uniformity across different datasets.

2. Attention to Detail:

- Enhanced attention to detail through careful data entry processes, minimizing mistakes and ensuring high-quality data.
- Improved ability to detect and correct discrepancies or anomalies during data entry.

3. Efficient Data Management:

- Gained experience in managing large volumes of data efficiently, organizing data in a structured manner for easy access and analysis.
- Learned to prioritize data entry tasks to meet deadlines without compromising accuracy.

4. Data Cleaning Techniques:

- Acquired skills in basic data cleaning techniques, such as handling missing values, standardizing data formats, and detecting duplicates.
- Understood the importance of clean data for reliable analysis and decision-making.

5. Use of Data Entry Tools and Software:

- Gained proficiency in using various data entry tools and software, enhancing overall efficiency and productivity.
- Learned to utilize features of data entry software to streamline the data entry process and ensure data integrity.

6. Manual Data Entry Skills:

- Improved manual data entry skills, ensuring precise and accurate entry of data that could not be automated.
- Enhanced ability to verify data against source documents to ensure correctness.

7. Data Verification and Validation:

- Developed the ability to verify and validate data to ensure it meets the required standards and criteria.
- Learned to cross-check data entries for accuracy and completeness, reducing the risk of errors.

8. Time Management:

- Improved time management skills by balancing speed and accuracy in data entry tasks.
- Learned to manage multiple data entry projects simultaneously, ensuring timely completion.

9. Handling Sensitive Information:

- Understood the importance of handling sensitive and confidential information with care and discretion.
- Learned to follow data privacy and security protocols to protect sensitive data.

10. Collaboration and Communication:

- Improved communication skills by effectively reporting data entry issues and discrepancies to supervisors.
- Enhanced collaboration skills by working with team members to ensure accurate data collection and entry.

11. Adaptability:

- Developed the ability to adapt to different data entry tasks and requirements, demonstrating flexibility and versatility.
- Learned to quickly familiarize with new data entry tools and processes.

12. Problem-Solving:

- Improved problem-solving skills by identifying and addressing data entry issues and inconsistencies.
- Learned to troubleshoot common data entry problems and implement effective solutions.

13. Data Integrity:

- Gained a deep understanding of the importance of data integrity and the role of accurate data entry in maintaining it.
- Learned to implement best practices to ensure the reliability and trustworthiness of data.

14. Analytical Skills:

- Developed basic analytical skills by reviewing and interpreting data entries to ensure they meet predefined criteria.
- Enhanced ability to identify patterns and trends within the data during the entry process.

15. Project Management:

- Gained experience in managing data entry projects from start to finish, ensuring all tasks are completed accurately and on time.
- Learned to create and follow detailed data entry plans and schedules.

16. Real-World Business Context:

- Understood the significance of accurate data entry in real-world business contexts, supporting decision-making and operational efficiency.
- Gained insights into how data entry contributes to larger data management and analysis projects.

Data analysis

1. Data Cleaning and Preparation:

- Gained proficiency in data cleaning techniques, ensuring data integrity by handling missing values, standardizing data entries, detecting and treating outliers, and transforming data into suitable formats for analysis.

2. Data Collection Skills:

- Developed skills in web scraping to gather large volumes of data from various sources using Python scripts and tools like Beautiful Soup and Scrapy.
- Enhanced manual data entry skills, ensuring accuracy and consistency in data collection.
- Conducted secondary research to gather additional contextual information, enriching the dataset.

3. Exploratory Data Analysis (EDA):

- Gained expertise in performing EDA using Power BI, uncovering underlying patterns, trends, and relationships within the data.
- Developed the ability to analyze various variables such as movie titles, release dates, genres, ratings, production houses, box office figures, and IMDB ratings to derive meaningful insights.

4. Insight Generation:

- Learned to generate actionable insights from data analysis, contributing to strategic decision-making processes.
- Gained practical insights into the dynamics of the film industry, understanding factors contributing to a movie's success or critical reception.

5. Enhanced Technical Skills:

- Improved proficiency in using data analysis tools and techniques, particularly in Power BI and Python.
- Gained experience in data cleaning, preparation, and analysis, enhancing overall technical skills.

6. Practical Application of Data Analytics:

- Understood the transformative potential of data analytics in driving business outcomes, particularly within the entertainment sector.
- Applied data analytics techniques to real-world business problems, contributing to operational efficiency and strategic decision-making.

7. Team Collaboration and Communication:

- Improved collaboration skills by working with a team to gather, clean, and analyze data.
- Enhanced communication skills by presenting findings and insights to stakeholders, ensuring clarity and understanding.

8. Problem-Solving and Critical Thinking:

- Developed problem-solving skills by addressing data inconsistencies, missing values, and outliers.
- Enhanced critical thinking abilities by evaluating the impact of various factors on the overall dataset and making informed decisions.

9. Project Management Skills:

- Gained experience in managing data analysis projects, ensuring timely completion of tasks and meeting project objectives.

10. Real-World Business Context:

- Applied theoretical knowledge to practical business scenarios, gaining a deeper understanding of how data analytics can be used to drive business success.

Their internship at Fipchain Technology Pvt Ltd provided a comprehensive platform to immerse themselves in the field of data analysis, allowing them to gain hands-on experience with real-world datasets and explore the practical applications of data-driven strategies in business contexts. The focal point of their work was a detailed project centered around a movie dataset, where they applied advanced techniques in Exploratory Data Analysis (EDA) to uncover valuable insights. This project involved cleaning and preparing the dataset, followed by rigorous analysis to identify patterns, trends, and relationships within the data. The use of tools such as Jupyter Notebooks, Power BI, and Excel was integral to this process, enabling them to apply statistical methods effectively and visualize the results in a meaningful way.

REFERENCES / BIBLIOGRAPHY

- Fipchain. (n.d.). Fipchain Technologies Pvt Ltd. LinkedIn. Retrieved August 23, 2024, from *https://www.linkedin.com/company/fipchain/*
- Company Check. (n.d.). Fipchain Technology Private Limited. Retrieved August 23, 2024, from *https://www.thecompanycheck.com/company/fipchain-technology-private-limited/U72200TN2021PTC146604*
- Producer Bazaar. (n.d.). Producer Bazaar. Retrieved August 23, 2024, from *https://producerbazaar.com/*
- Medianews4u. (n.d.). Fipchain Technology Pvt Ltd. Retrieved August 23, 2024, from *https://www.medianews4u.com/tag/fipchain-technology-pvt-ltd/*