

SUMMER INTERNSHIP PROGRAM

AT

UNIMRKT RESPONSE LLP



FROM 12th May 2025 to 31st July 2025

A REPORT BY

MUSKAN SINGH

Master of Business Administration
(HEALTHCARE ANALYTICS)

2024-2026

Under the mentorship of

Dr. Ritu Vashistha
Assistant Professor
School of Digital health
IIHMR, Jaipur



JAIPUR



Internship Completion Certificate

Dated: - 1st August 2025

This is to certify that **Muskan Singh**, D/O **Mrs. Shanti Kumari Singh** has successfully completed their summer internship at **Unimrkt Response LLP** from **12th May 2025 to 31st July 2025**.

During her tenure, she contributed to various live projects and demonstrated strong technical skills in:

- **Python & SQL** – Developed data pipelines and performed data analysis for large-scale datasets.
- **Power BI Dashboards** – Designed and deployed dynamic dashboards for real-time business insights.
- **Elasticsearch** – Queried and indexed large datasets to support fast and efficient data retrieval.
- **Kibana Dashboards** – Built and customized dashboards to visualize operational and user data.
- **Live Dashboards** – Delivered interactive, auto-refreshing dashboards for continuous monitoring and reporting.

Ms. Muskan displayed excellent problem-solving ability, professionalism, and a strong grasp of accounting tools. Her work added real value to our finance processes.

We wish **Muskan Singh** all the best in her future endeavours.

Warm regards,



Shiva Tawar
Senior Human Resources Executive
Unimrkt Response LLP

IIHMR University Faculty Mentor Approval

This is to certify that **Ms. Muskan Singh** of academic year 2024-26 **School of Digital Health** has prepared a poster with respect to her summer internship held during May-July 2025.

She has carried out work as per the guidelines. Her poster is approved for presentation.

Date: 05.08.2025

A handwritten signature in blue ink, appearing to read 'Ritu'.

(Signature of Faculty Mentor)

Name: Dr. Ritu Vashistha

Designation: Assistant Professor

WEEK 1

Name of the Organisation: UNIMRKT RESPONSE LLP

Area/ Department/ Project of Summer Internship Program: Data Analytics

Name of the Student: Muskan Singh

Roll no: 0016

Stream: Healthcare Analytics

Week no: 01

From: 13/05/2025 – 16/05/2025

Activity Done	- I established a strong background in SQL during the internship by exercising basic commands such as SELECT, WHERE, GROUP BY, ORDER BY, and filtering data. I also learned to apply various types of joins such as INNER, LEFT, RIGHT, and FULL JOINS, and applied these in combining data from various tables in SQL Server. - A significant portion of my assignment was doing manual data entry for real-time updates, which I then used to build simple dashboards in Power BI. From this, I experimented with primary visualization tools such as charts, filters, and slicers.
Linking theory (Classroom learning) with practice at your organisation	I practiced classroom learning in an actual scenario—particularly the significance of clean, organized data to proper reporting.
Gaps identified on the working area.	Manual entry of data was labor-intensive and error-prone, and I originally found it difficult to implement sophisticated joins with larger datasets. I also had some doubt when handling live data consistency between departments and required more understanding of how Power BI interacts with SQL databases.
Suggestions for improvement	- Arrange a short training on advanced SQL queries and best practices. - Provide clear documentation or workflow for live data handling.
Plan for the next week	- Learn more of advanced SQL queries including subqueries, CTEs, and window functions. - Practice more data cleaning and transformation techniques in Power BI.

Signature of the Student

Muskan Singh

WEEK 2

Name of the Organisation: Unimrkt Response LLP

Area/ Department/ Project of Summer Internship Program: Data Analytics

Name of the Student: Muskan Singh

Roll no: 0016

Stream: Healthcare Analytics

Week no. 02

From: 19/05/2025 – 23/05/2023

Activity Done	• In this stage of my internship, I spent a lot of time working with a market research dataset covering suppliers, surveys, and clients that was made available by the company. • I also continued to gain first-hand experience with SQL Server, where I experimented with advanced SQL features such as window functions (such as RANK, ROW_NUMBER, LEAD, and LAG), and with the usefulness of stored procedures in constructing reusable blocks of queries. Although I achieved a general insight into indexing and how it affects query performance, I understood that this is an area requiring more thorough learning on my end.
Linking theory (Classroom learning) with practice at your organisation	• I learned the usefulness of stored procedures in controlling repeated queries and solidified my Excel skills by doing actual company projects, which connected my learning to real-world business requirements.
Gaps identified on the working area.	• Difficulty in grasping indexing in depth and its impact on performance. • Limited hands-on guidance for forecasting accuracy in Excel.
Suggestions for improvement	• Give more guided forecasting assignments using varied business scenarios. • Provide reference materials for Excel What-If tools and forecast functions.
Plan for the next week	• Looking ahead, I think it would be beneficial to continue doing more guided forecasting exercises through different business scenarios, obtain reference materials that emphasize Excel forecasting tools, and receive greater exposure to SQL performance optimization techniques

Signature of the Student

Muskan Singh

WEEK 3

Name of the Organisation: Unimrkt Response LLP

Area/ Department/ Project of Summer Internship Program: Data Analytics

Name of the Student: Muskan Singh

Roll no: 0016

Stream: Healthcare Analytics

Week no: 03

From: 26/05/2025 – 30/05/2025

Activity Done	• This week, I emphasized solidifying my Python building blocks. I did all basics through to dictionaries and implemented concepts such as variables, loops, functions, and data structures in assignment-based problem-solving using Google Colab. I also got introduced to data visualization using Matplotlib.
Linking theory (Classroom learning) with practice at your organisation	• Applied foundational knowledge of Python to solve real-world problems. • Gained exposure to data visualization principles using Matplotlib from academic sessions.
Gaps identified on the working area.	• I honed logic construction, question development, and visual modeling through coding labs, although I struggled at first with Matplotlib syntax and nested dictionaries. These experiences pointed to needing more guided feedback and visual samples.
Suggestions for improvement	• Organize doubt-clearing sessions or feedback discussions on Python assignments. • Provide a cheat sheet or visual reference for Matplotlib chart types and code. • Share problem-solving strategies for query writing and logic development. • Encourage group discussions or peer review of Python code for better understanding.
Plan for the next week	• For further learning, I intend to work with real datasets using Pandas, learn more about Matplotlib, and start combining Python with tools such as SQL or Excel.

Signature of the Student-

Muskan Singh

WEEK 4

Name of the Organisation: Unimrkt response LLP

Area/ Department/ Project of Summer Internship Program: Data Analytics

Name of the Student: Muskan Singh

Roll no: 0016

Stream: Healthcare Analytics

Week no: 04

From: 02/05/2025 – 06/05/2025

Activity Done	➤ Learned Matplotlib in-depth, covering: • Line Plot • Bar Chart and Horizontal Bar Chart • Subplots • Step Plot • Pie Chart • Histogram • Other plots like scatter plot, area plot, etc. • Labels, Titles, Legends • Colors, Styles, Widths • Axes control and grid adjustments ➤ Completed an assignment on Matplotlib using various plots and properties to visualize different types of data.
Linking theory (Classroom learning) with practice at your organisation	• Applied data visualization concepts taught in class to real or mock data using Matplotlib which has not been yet taught in class. • I attained clarity on how to use the right kind of chart for which type of data.
Gaps identified on the working area.	• Initial difficulty in choosing appropriate plots for certain data types. • Needed more clarity on subplot grid configurations.
Suggestions for improvement	• Provide practice datasets to improve decision-making on plot types.

Plan for the next week	• Start working on a mini project integrating SQL, Python, and visualizations. • Explore real datasets from the organization for practice.
-------------------------------	--

Signature of the Student

Muskan Singh

WEEK 5

Name of the Organisation: UNIMRKT RESPONSE LLP

Area/ Department/ Project of Summer Internship Program: DATA ANALYTICS

Name of the Student: MUSKAN SINGH

Roll no: 0016

Stream: MBA HEALTHCARE ANALYTICS

Week no: 05

From: 9/05/2025 – 13/05/2025

Activity Done	• Started learning Pandas in Python. Completed 2 assignments on Pandas (based on company datasets). • Completed 1 SQL project and uploaded it on GitHub. • Created GitHub account. • Learned and practiced Seaborn visualizations (barplot, boxplot, countplot, etc.) • Uploaded all my dashboards, and visualizations of MATPLOTLIB AND SEABORN onto Github
Linking theory (Classroom learning) with practice at your organisation	• Applied Python and data analysis skills learned in class to real-time data • Used concepts of DataFrames, filtering, and plotting from classroom learning • Integrated Seaborn charts for visual presentation of business data
Gaps identified on the working area.	• Need more exposure to advanced Pandas functions (groupby, merge, pivot) • Lack of fluency in interpreting large datasets quickly • Required support in setting up GitHub repositories properly
Suggestions for improvement	• Spend more time with real datasets to build deeper insight • Participate in short GitHub practice tasks/tutorials • Revise and document the Pandas functions used during assignments
Plan for the next week	• Deep dive into advanced Pandas (groupby, merging, pivoting) • Start working on a mini data project using Python (Pandas + Matplotlib) • Learn basics of Logstash and Elasticsearch if applicable

Signature of the Student

Muskan Singh

WEEK 6

Name of the Organisation: Unimrkt Response LLP

Area/ Department/ Project of Summer Internship Program: Data Analytics

Name of the Student: Muskan Singh

Roll no: 0016

Stream: MBA HA 01

Week no: 06

From: 16/05/25 – 22/05/25

Activity Done	I finished completing 2 assignments based on data manipulation and visualization. I also practiced NumPy and Pandas operations and worked on creating charts using Matplotlib.
Linking theory (Classroom learning) with practice at your organisation	I conducted array operations, DataFrame manipulations, and basic data visualization in real world data.
Gaps identified on the working area.	It was difficult to understand the calculations and changes made at first but later on, I got a hang of it.
Suggestions for improvement	Need to focus more on practical implementation through mini projects. Consider revisiting official documentation and working on real datasets to improve speed and clarity.
Plan for the next week	Begin learning how to clean and preprocess messy datasets. Started a small project involving real-life CSV files to integrate everything learned so far.

Signature of the Student

Muskan Singh

WEEK 7

Name of the Organisation: Unimrkt Response LLP

Area/ Department/ Project of Summer Internship Program: Data Analytics

Name of the Student: Muskan Singh

Roll no: 0016 Stream: MBA Healthcare Analytics

Week no: 07

From: 23/06/2025 – 27/06/2025

Activity Done	Worked on a project that integrated NumPy, Pandas, and Matplotlib for data analysis and visualization. Also practiced and applied aggregate functions, window functions, conversion functions, and logical functions in SQL.
Linking theory (Classroom learning) with practice at your organisation	Applied theoretical knowledge of Python libraries and SQL functions to solve real data problems, manipulate datasets, and create visualizations.
Gaps identified on the working area.	SQL Query optimization should be undertaken
Suggestions for improvement	I need to explore larger datasets for deeper hands-on learning and extensive practice.
Plan for the next week	I need to work on the integration of Power BI and SQL along with Python.

Signature of the Student

Muskan Singh

WEEK 8

Name of the Organisation: Unimrkt Response LLP

Area/ Department/ Project of Summer Internship Program: Data Analysis

Name of the Student: Muskan Singh

Roll no: 0016 Stream: MBA HA 01

Week no: 08

From: 30-06-2025 to 04-07-2025

Activity Done	• I began my Inpatient Healthcare Analysis project. I investigated the dataset with 34 columns and 2.7 million rows. I loaded and inspected the data with pandas, determined data types, and verified for nulls, duplicates, and inconsistencies. • I started cleaning by eliminating duplicates and filling in missing values. Also, I performed preliminary calculations with Numpy to gain an understanding of general statistics (e.g., average, median, and distribution of important variables such as charges and length of stay).
Linking theory (Classroom learning) with practice at your organisation	• Worked and practiced using the principles of data cleaning, exploratory data analysis, and Python packages (pandas, NumPy) taught in class to process big data effectively.
Gaps identified on the working area.	• There were high missing values in some of the columns (e.g., secondary diagnosis codes, discharge disposition). There were columns that were poorly named or had inconsistent formatting.
Suggestions for improvement	• Data had a lot of inconsistencies.
Plan for the next week	• I need to complete the visualization part using group by and other aggregate functions. Also, I have to start making the dashboard. Moreover, I also have to work on understanding Logstash- a part of ELK stack

**Signature of the Student-
Muskan Singh**

WEEK 9

Name of the Organisation: Unimrkt Response LLP

Area/ Department/ Project of Summer Internship Program: Data Analysis

Name of the Student: Muskan Singh

Roll no: 0016

Stream: MBA Healthcare Analytics

Week no: 09

From: 7/07/2025 – 11/07/2025

Activity Done	I worked extensively on a data analysis project using Python. Initially, I made an error while cleaning and analyzing the data, which required me to redo the process from the beginning. After correcting the blunder, I successfully visualized the data and started working on creating an interactive dashboard.
Linking theory (Classroom learning) with practice at your organisation	I applied concepts from our Data Analytics and Python programming, including the use of libraries such as - Pandas, - NumPy, - Seaborn, and Matplotlib. The process of data wrangling, visualization, and dashboard creation connected directly with what I've learned in Healthcare Analytics and Business Intelligence modules.
Gaps identified on the working area.	Data inconsistencies could have been checked
Suggestions for improvement	To avoid rework, I suggest implementing a review checklist before starting visualization — especially around data types and missing value handling.
Plan for the next week	- I plan to finish building the dashboard with filters, dropdowns, and clean visualization layers - Also, I will be learning Logstash.

Signature of the Student

Muskan Singh

WEEK 10

Name of the Organisation: Unimrkt Response LLP

Area/ Department/ Project of Summer Internship Program: Data Analytics

Name of the Student: Muskan Singh

Roll no: 0016

Stream: MBA Healthcare Analytics

Week no: 10

From: 14.07.2025 – 18.07.2025

Activity Done	• I completed the analysis of "UNCOVERING PATTERNS IN INPATIENT HOSPITAL DATA: A STATISTICAL AND VISUAL STUDY" this week utilizing a number of statistical techniques to determine trends in cost, types of admissions, and geographic differences. Visualizations were enhanced using Power BI, and conclusions were drawn to the final stage for presentation. • I commenced work on my second project titled "Behavioral and Demographic Insights from Survey Response Patterns: A Data-Driven Evaluation" • I created "Measures" and dealt with questions to be responded to and set up the visualizations.
Linking theory (Classroom learning) with practice at your organisation	Key concepts from our Statistics and Data Visualization courses were utilized such as descriptive statistics, and outlier detection. Employing SQL and Excel to clean data and Power BI for dashboards aligned directly with academic coursework regarding healthcare analytics.
Gaps identified on the working area.	When using actual world data, I noticed missing data entries and uneven formatting in some records, which affected precision. Additionally, there was minimal automation when dealing with big data, which made the original processing slow.
Suggestions for improvement	Automation tools like Python or R could be introduced
Plan for the next week	Incorporate mentor feedback and prepare for the final project presentation.

Signature of the Student
Muskan Singh

WEEK 11

Name of the Organisation: Unimrkt Response LLP

Area/ Department/ Project of Summer Internship Program: Data Analytics

Name of the Student: Muskan Singh

Roll no: 0016 Stream: MBA HA

Week no: 11

From: 21.05.2025 to 25.05.2025

Activity Done	• SQL Optimization • DAX Calculations in Power BI • Dashboard enhancement for executive reporting
Linking theory (Classroom learning) with practice at your organisation	Applied SQL query optimization techniques and DAX functions learned during coursework to real datasets and dashboards used for business reporting.
Gaps identified on the working area.	Minor issues with data refresh lag and optimizing DAX for performance in large datasets.
Suggestions for improvement	Nothing
Plan for the next week	Work on Logstash and Data optimization

Signature of the Student

Muskan Singh

WEEK 12

Name of the Organisation: Unimrkt Response LLP

Area/ Department/ Project of Summer Internship Program: Data Analytics

Name of the Student: Muskan Singh

Roll no: 0016 Stream: MBA Healthcare Analytics

Week no: 12

From: 28.07.2025 to 1.08.2025

Activity Done	• Learning the basics of Logstash (ELK Stack) • Learning from Udemy • Completed a course on the same • Practical experience gained
Linking theory (Classroom learning) with practice at your organisation	• This is not a part of my syllabus
Gaps identified on the working area.	• None
Suggestions for improvement	• I have just started working on this. Need some more time to completely familiarize with it
Plan for the next week	• My internship ends on the 1st of August

Signature of the Student

Muskan Singh

COMPILED REPORT

Name of the Organisation: Unimrkt Response LLP

Area/ Department/ Project of Summer Internship Program: Data Analytics

Name of the Student: Muskan Singh

Roll no: 0016 Stream: MBA HA-01

Week no: 12 weeks

From: 12th May 2025 – 1st August 2025

Activity Done	• I started off with Power BI and Tableau visualization • Excel Worked on-advanced formulas, Macros, What If Analysis and Forecasting • Python and SQL-join, query optimization, nested queries, indexing and slicing. Numpy, Pandas, Matplotlib, and Seaborn. • In addition, Kibana visualization, basics of ML- supervised, unsupervised, and reinforcement learning, and Logstash-a part of ELK stack.
Linking theory (Classroom learning) with practice at your organisation	• In class I had already learned a part of tableau and PBI but during my internship period I learnt SQL and Python in depth and detail. Also, my project is based on the same.
Gaps identified on the working area.	• I can work more on learning advanced queries, practicing nested queries and working on projects.
Suggestions for improvement	• Practice is the key to being Perfect
Plan for the next week	• Only Practice.

Signature of the Student-

Muskan Singh

PROJECT 1: INPATIENT HOSPITAL ANALYSIS

Introduction:

1. Background of the Study

Hospital charges in the healthcare industry vary substantially by region and patient population. Knowledge about what drives these costs is important to stakeholders like policymakers, healthcare providers, and insurers.

2. Importance of Cost Transparency

Increased healthcare costs create a need for greater transparency in hospital billing. Charge data analysis informs discrepancies and inefficiencies in the healthcare system.

3. Purpose of the Project

The objective of this project is to examine the total hospital charges corresponding to different admission types, risk categories, and sites, thus providing information on the economic costs of health care services.

4. Data-Driven Decision Making

The results of this research are meant to inform data-driven decisions for allocating resources, setting prices, and policy formulation.

5. Scope of Analysis

Analysis is centered around publicly accessible hospital charge data, looking at trends by admission types, risk of mortality, and by geographic areas like urban and rural counties.

Objectives:

1. To Analyze Hospital Charges Across Admission Types

Learn how emergency, urgent, elective, trauma, and newborn admissions vary in terms of overall charges.

2. To Explore the Impact of Risk of Mortality

Investigate the impact of the severity of a patient's condition (minor, moderate, major, extreme) on overall hospital billing.

3. To Investigate Geographic Variability

Determine variations in hospital charges across urban versus rural counties, with particular attention to high-cost districts such as Manhattan, Queens, and Kings.

4. To Identify Patterns and Anomalies

Identify any patterns, outliers, or surprising spikes in data that need further examination or action from healthcare organizations.

5. To Develop Visual Insights

Develop effective visualizations that enable non-technical stakeholders to understand significant findings and trends in hospital expenditures.

Methodology:

1.Data Collection

Historical charge data from hospitals was obtained from an open government or health portal. Variables such as total charges, type of admission, risk of mortality, county of the hospital, and patient demographics were included.

2.Data Cleaning and Preprocessing

The data were cleaned with Python libraries like Pandas. The missing values were handled, and irrelevant or redundant entries were eliminated for ensuring data integrity.

3. Data Segmentation

The data were filtered out according to certain attributes like 'APR Risk of Mortality', 'Admission Type', and 'Hospital County' to identify segments of interest for analysis.

4. Descriptive and Graphical Analysis

Seaborn and Matplotlib were utilized to create visual insights like bar plots, histograms, and trend graphs. Descriptive statistics including mean, median, and standard deviation were computed.

5.Insights and Interpretation

The statistical and visual outputs were interpreted to conclude relevant facts regarding hospital charge trends. Particular emphasis was given to trends that reflected regional or categorical differences in billing.

Tools Used:

1. Python (Pandas, NumPy):

It was used for data cleaning, manipulation, and preprocessing. These libraries helped in filtering, aggregating, and segmenting the dataset based on attributes such as admission type, mortality risk, and hospital county.

2. Matplotlib and Seaborn:

These data visualization libraries were used to create a wide range of plots such as bar graphs, histograms, violin plots, factor plot, and boxplots.

3. Jupyter Notebook:

It provided a platform to write codes and run it as well which helped analyze and understand the patterns present in the dataset.

4. MS Excel (Preliminary Review):

Initial exploration and basic descriptive summaries were conducted before embarking upon Python and its libraries.

5. Google Colab (Optional Tool):

A lot of practice was done on various sheets that was shared over this platform upon which I started working on the actual interface.

Skills Applied:

- 1. Data Cleaning & Preprocessing:**
Applied techniques to handle missing values, remove duplicates, and transform inconsistent data to ensure the dataset was ready. I also looked out for blanks and null values and arranged the data in ascending or descending order.
- 2. Exploratory Data Analysis (EDA):**
I conducted EDA meticulously to uncover relationships between various factors such as admission types, geographic distribution, and mortality risks in relation to hospital charges.
- 3. Statistical Interpretation:**
I tried interpreting various statistical measures such as Mean, median, and Mode which was used to understand the statistics related to it.
- 4. Data Visualization:**
the Python libraries helped me transform raw numbers into understandable visual insights that highlighted charge disparities across counties, risk categories, and types of care.
- 5. Analytical Thinking and Insight Generation:**
Employed domain reasoning to make sound conclusions from the findings and interpret them into sensible and meaningful insights applicable to healthcare stakeholders as well as policy consideration.

Findings:

- 1. Emergency and Urgent Admissions Incur the Highest Charges**
Emergency and urgent admission categories always had the highest hospital charges in all counties. They also exhibited the highest variability, suggesting that unanticipated and high-acuity cases significantly contribute to costs.
- 2. High Risk of Mortality Correlates with Higher Costs**
Patients in the "Major" and "Extreme" risk of death categories had significantly higher charges than those in "Minor" or "Moderate" risk categories. This is evidence that the severity and complexity of care for critically ill patients are large cost factors.
- 3. High-Cost Incidents Disproportionately Occur In Urban Counties**
Manhattan, Queens, and Kings counties recorded the highest density of expensive admissions. Urban cities probably have higher fees from a combination of influential factors such as sophisticated healthcare facilities, specialized care, and urban cost inflation.
- 4. Discrepancy Between Admission Types and Charge Uniformity**
Although trauma and electives were included in the database, these were found with relatively lower and stable hospital charges. This can be due to the reason that these admissions are either planned or controlled, where utilization can be expected.
- 5. Geographic and Demographic Patterns Frames Cost Landscape**
The study found a high correlation between geography, urgency of admission, and severity. Urban areas with a high incidence of extreme cases also had higher charges, implying the necessity for research on underlying socioeconomic or population health determinants.

Conclusion:

1. Significant Variability Continues to Persist in Hospital Charges

The research establishes that hospital prices are not standard and are highly variable by admission type, illness severity, and location. These differences are highlighted by the sheer complexity of healthcare pricing systems. These disparities underscore the complexity of healthcare pricing mechanisms.

2. Critical Care and Emergencies Surge Costs

Emergency care and mortal risk cases substantially increase aggregate costs. These results are indicative that stronger preventive care and early intervention could lower such high-cost cases.

3. Urban Centers Bear the Financial Brunt

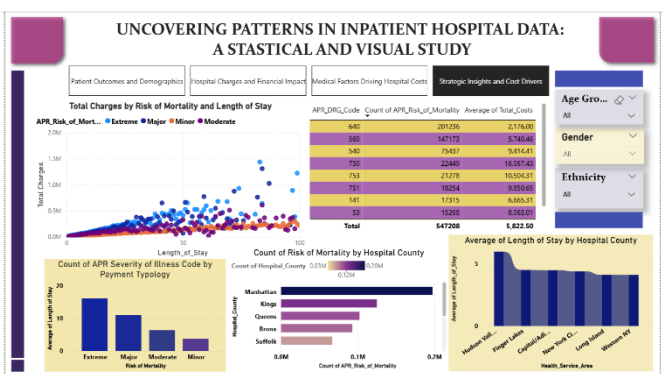
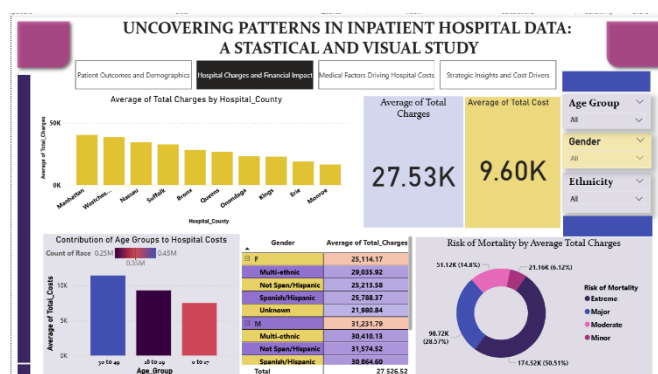
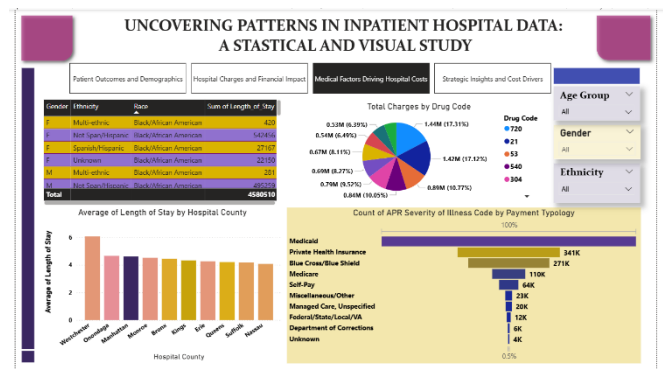
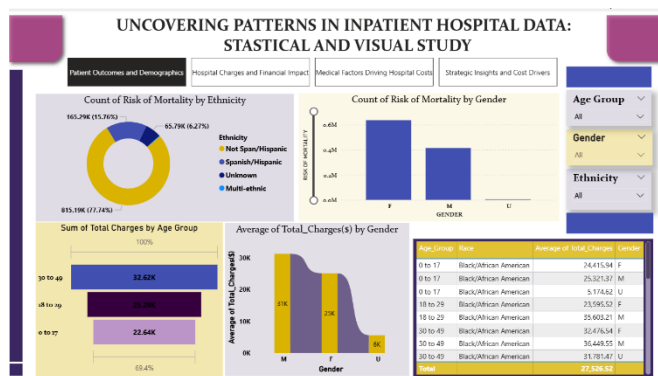
Urban counties appear as both sources and clusters of high-charge hospitalizations. This is evidence of the double realities of increased availability of care and higher costs of services within urban healthcare systems.

4. Charge Patterns Can Guide Policy and Planning

The identified patterns can educate healthcare planners, policy makers, and insurers on cost concentrations. This empirically derived information is crucial for fair pricing reforms and effective resource allocation.

5. Need for Greater Transparency and Regional Equity

The analysis promotes equitable, transparent, and regulated pricing schemes and enhanced accessibility of care between counties. Closing urban-rural cost disparities and enhancing risk management options can result in more balanced and sustainable healthcare provision.



PROJECT 2: Behavioural and Demographic Insights from Survey Response Patterns A Data-Driven Evaluation

Introduction:

This project involves analyzing response data from surveys in order to reveal important patterns, behavior, and cost associated with survey completion. It seeks to offer a strategic snapshot of survey characteristics, participant behavior, completion trends, and the financial efficiency of survey execution. The project produces business-ready dashboards to enable stakeholders to make informed decisions on survey targeting, budgeting, and optimization.

Objectives:

- Analyze respondent behavior (completes vs. incompletes).
- Evaluate and analyze cost and revenue trends per country, client, and supplier.
- Understand country/language-specific trends in survey participation.
- Investigate average Length of Interview (LOI) patterns across devices and buyers.
- Derive actionable business conclusions on supplier performance and respondent quality.

Methodology:

- Data was imported into Power BI and cleaned using Power Query to handle inconsistencies, null values, blanks, and formatting issues.
- Relevant columns were transformed.
- Data types were corrected to ensure analytical accuracy.
- Custom columns and Conditional columns were worked upon.
- Custom measures and calculated columns were created using DAX to compute metrics like CPI, completion rate, and total cost.
- Visualizations such as bar charts, pie charts, cards, and metric tables were accommodated. Mostly bar charts were used to ensure clarity of insight presentation.

Tools Used:

- **Power BI:** For data cleaning, transformation, DAX calculations, and dashboard development.
- **DAX (Data Analysis Expressions):** For creating calculated measures (e.g., average CPC, completion rate, and derived metrics).
- **Power Query Editor:** For preprocessing the raw data and performing cleaning operations.

Skills Applied:

- Data cleaning and transformation.
- Data modeling and relationships set up.
- Basic and intermediate DAX functions applied (e.g., SUM, AVERAGE, DIVIDE, IF conditions)
- Visualization design and storytelling with multiple dashboards
- Finalizing the dashboard to showcase the important metrics.
- Performance metrics interpretation and insight generation.

Key Findings-

➤ Survey Completion Overview

- Total Completes: 15K, Total Incompletes: 235K
- Major reasons for incompletes: Client Terminate, Bad User, Client Redirect
- Suppliers with highest total responses: Dynata Router, Toluna, Zampla
- Most incomplete surveys are concentrated around a few specific survey IDs.

➤ Cost & Revenue Overview

- Total Cost: ₹276.71K, Total Revenue: ₹385.22K
- Highest cost and revenue country: US
- Most costly supplier per full: Purespectrum for ₹76.26 per survey
- Lowest CPI supplier: Quick Rewards, Cint API
- Link, Dynata Router, and SampleCube MBD API were leading buyers by total cost.

➤ Country & Language Insights

- Greatest completes from US (7.2K), followed by GB (1.1K) and AU (0.8K)
- Most frequent survey languages: Arabic, Spanish, English
- Countries with highest CPI: Argentina (₹5.67), Saudi Arabia, and UA
- Romania was high cost and low revenue – potentially loss-making.

➤ Survey Characteristics (LOI Insights)

- LOI (Length of Interview) varies between 2-80 minutes with an average of 11.72 minutes
- Device with most completes: Device 2
- Supplier with highest LOI: RevenueClick Media (25 mins)
- Buyer with highest average LOI: The Paradigm Shift – Service (59 mins)

Conclusions:

- **Optimization Required:** An extremely high percentage of incomplete responses (above 94%) suggests inefficiency in targeting, quality, or routing.
- **Top Performers Detected:** Some suppliers such as Dynata and Toluna provide the majority of responses; However, quality is not consistent.
- **Geographic Insights:** US is the most efficient region both in terms of completes and revenue. Romanian and Bulgarian cost-efficiency is bad.
- **LOI Management:** Higher LOIs have a larger number of drop-offs. Buyers and suppliers with very high LOIs should be reviewed.

- ## DASHBOARDS

