

Project Title :
Visualization of healthcare data in a specific geographical region
(Case study-Bihar)

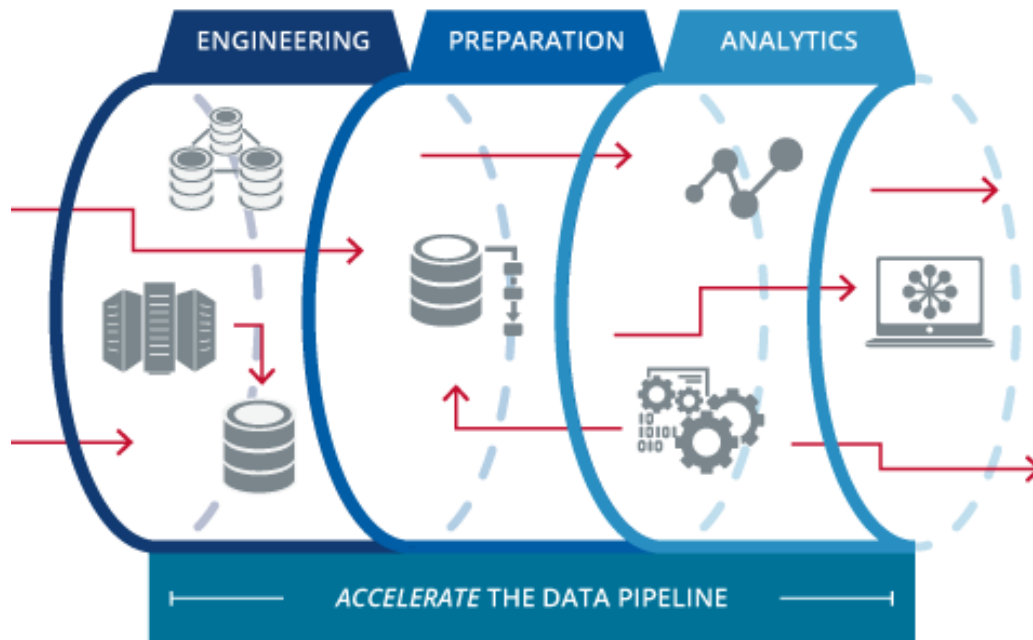
INTERNSHIP PROJECT REPORT

AIM : To create an interactive dashboard that modernizes and provides a new outlook of the dataset present.

PRINCIPLE :

- Data should be our main concern, along with finding a way to transform it into actionable insights. Nowadays, data is considered a high value resource to any company seeking to lead in the rapidly evolving world.
- Efficient data transportation relies on three important blocks:
 1. Data Producers: Data source points where raw data lies ready to be fetched.
 2. Transformation and Transportation workflows: ETL sub processes that involve stacks of *extraction, transformation and data loading* layers to route the data to its corresponding end points.
 3. Data Consumers: Final end points that utilize clean and preprocessed information to perform high end tasks.
 4. ETL SCHEMA: (EXTRACT, TRANSFORM, LOAD)

Project Title :
Visualization of healthcare data in a specific geographical region
(Case study-Bihar)



THE LINK BETWEEN SQL AND MACHINE LEARNING

- SQL is needed for machine learning. It is the de facto standard language for querying data; it is required to format data to be used by machine learning algorithms for improved pattern detection.
- The link between machine learning and SQL is data. Processing the amount of data required for machine learning requires proper querying. SQL is the language of choice to query data.
- SQL allows data scientists and machine learning engineers to obtain the raw material for machine learning data.
- SQL rises above Python or R not because it is more powerful or more robust. Its importance comes from being the language used for “first contact” with the data needed for machine learning. SQL is the most straightforward language to query

Project Title :

Visualization of healthcare data in a specific geographical region
(Case study-Bihar)

data. Additionally, in learning SQL, you can leverage efficiencies later by using SQL and Python in tandem.

- Another use for SQL is to run and build learning models for machine learning inside a database. In other words, conduct data queries, perform data analysis and run algorithms without performing a fetch function that has to pull data to an outside platform. Everything is done literally within the database.

In a machine learning workflow, data is king. The more relevant data you have, the better your modeling and pattern detection will be.

SCOPE OF THE PROJECT:

- Microsoft SQL Server Machine Learning Services allows us to execute Python and R scripts in-database. We used it to prepare and clean data, do feature engineering, and train, evaluate, and deploy machine learning models within a database. The feature runs our scripts where the data resides and eliminates transfer of the data across the network to another server.
- To create a dashboard that would help in Modern data visualization and convert complex data into user-friendly visuals that are easy to understand for its stakeholders, doctors, patients, or government officers.
- To provide a new and improved dashboard that contains mainly the DHC(District health Centre) and PHC(Primary health centre) wise details of the state of Bihar.
- Make better decisions which is done by unifying diverse medical data coming from the dataset.

PROPOSED SOLUTION:

Our model consists of 7 steps:

1) Get Data: In this step, we are going to import data from the Microsoft SQL database.

Project Title :

Visualization of healthcare data in a specific geographical region
(Case study-Bihar)

2) **Fetch**: Selecting the required tables namely- patient count,PHC wise details, DHC wise details.

3) **Process**: In this step, the data is truncated and edited using several operations while loading the data into the Power BI.

4) **Analyze**:To analyze the processed data, and provide a new look.

5) **Visualize**: In this step, the analyzed and processed data is visualized through means of visuals such as Microsoft Power BI Visuals.

6) **Editing**: Rectifying errors.

7) **Web**: It is a state where the report is converted into the dashboard, and it can be shareable via URLs, websites.

BENEFITS OF THE SOLUTION:

- **Improved patient care**: By visualizing real-time data on patient health status, doctors can define and group patients according to the treatment and attention they need. Including all patient records in a dashboard also prevents practitioners from missing important information.
- **Trends and pattern recognition**: By using data visualization software that is often paired with built-in predictive analytics tools, doctors can estimate changes in patient health status and make better diagnoses.
- **Data presentations for different audiences**: Visuals can help a lot when you need to present data to business owners and other stakeholders or use it in healthcare marketing campaigns.
- **Accelerated performance**: Unlike lengthy, manual reporting, real-time data visualization significantly speeds up the analysis of information, helping healthcare organizations minimize process

Project Title :

Visualization of healthcare data in a specific geographical region
(Case study-Bihar)

inefficiencies, accelerate the decision-making process, and optimize costs.

- **Errors and fraud detections:** Data visualization improves the transparency of connections between patients, practitioners, providers, payers, and claims.

REFERENCES/SOURCES:

- **Analysis and Design of Visualization of Educational Institution Database using Power BI Tool** By Mandava Geetha Bhargava, K. Tara Phani Surya Kiran & Duvvada Rajeswara Rao
- **Interactive Visualization of Healthcare Data Using Tableau** By Inseok Ko, MS , Hyejung Chang, PhD
- **Healthcare Data Visualization: Geospatial and Temporal Integration** By Shenhui Jiang , Shiaofen Fang , Sam Bloomquist, Jeremy Keiper , Mathew Palakal, Yuni Xia and Shaun Grannis
- **Kshema - A Unified Healthcare Management Solution For Improving Efficiency of the Healthcare Delivery System in Rural India** By Anant R Koppar

POWER BI IN MACHINE LEARNING :

- Automated Machine Learning (AutoML) in Power BI for dataflows provides data scientists with the tools needed to train, validate and deploy Machine Learning (ML) models within Power BI.
- It takes Power BI's natural language processing capabilities to a different level to more simple applications like the Q&A visual.
- To use machine learning in Power BI, we simply create a dataflow and select the training data we would like to use.
- AutoML will then automatically extract the most relevant features and select a model that best fits the data. Finally, it will tune and validate the ML Model based on your desired outcome.

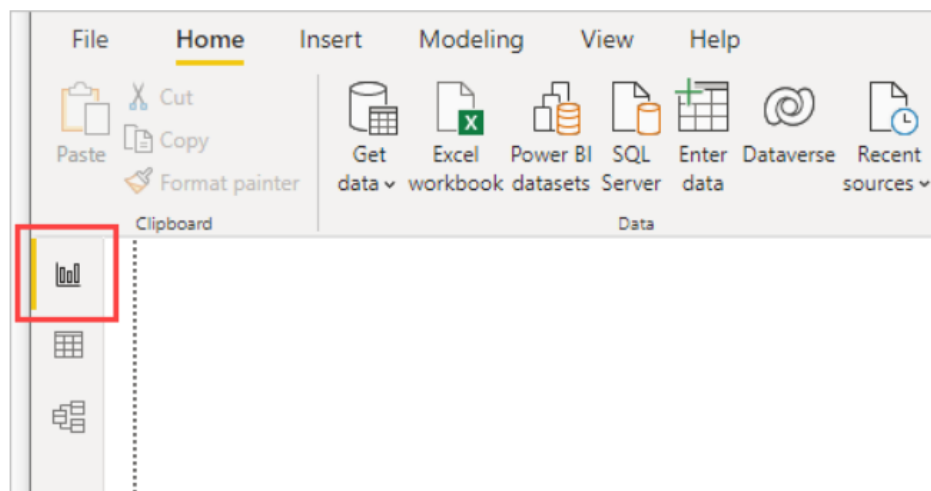
Project Title :
Visualization of healthcare data in a specific geographical region
(Case study-Bihar)

POWER BI DESKTOP :

- With Power BI Desktop we can connect to the world of data, create compelling and foundational reports, and share our efforts with others – who can then build on our work, and expand their business intelligence efforts.

Power BI Desktop has three views:

- Report view – where we use queries to create, to build compelling visualizations, arranged as we want them to appear, and with multiple pages, that we can share with others.
- Data view – see the data in our report in data model format, where we can add measures, create new columns, and manage relationships
- Relationships view – get a graphical representation of the relationships that have been established in our data model, and manage or modify them as needed.



- Power BI Desktop also comes with Power Query Editor.

Project Title : Visualization of healthcare data in a specific geographical region (Case study-Bihar)

- We use Power Query Editor to connect to one or many data sources, shape and transform the data to meet our needs, then load that model into Power BI Desktop.
- Here's how Power Query Editor appears once a data connection is established:
 1. In the ribbon, many buttons are now active to interact with the data in the query.
 2. In the left pane, queries are listed and available for selection, viewing, and shaping.
 3. In the center pane, data from the selected query is displayed and available for shaping.
 4. The Query Settings pane appears, listing the query's properties and applied steps.

The screenshot displays the Power Query Editor window titled "hospital records - Power Query Editor". The ribbon at the top includes tabs for File, Home, Transform, Add Column, View, Tools, and Help. The Transform tab is active, showing various data manipulation options like Close & Apply, New Source, Recent Sources, Enter Data, Data source settings, Manage Parameters, Refresh Preview, Advanced Editor, Choose Columns, Remove Columns, Keep Rows, Remove Rows, Sort, Split Column, Group By, Data Type: Text, Use First Row as Headers, Replace Values, Merge Queries, Append Queries, Combine Files, Text Analytics, Vision, and Azure Machine Learning.

On the left, the "Queries [9]" pane lists several queries, with "VW_PATIENTCOUNT" selected. The central pane displays a table with the following columns: ALC_DHCName, ALC_PHCName, ALC_SHCNAME, i23_patientcount, ALC_ReportFooter, ALC_PHCCode, and RegDate. The table contains 28 rows of data, including locations like Araria, Bhojpur, Begusarai, and Champaran-East.

On the right, the "Query Settings" pane is open, showing the "PROPERTIES" section with the query name "VW_PATIENTCOUNT" and the "APPLIED STEPS" section, which lists "Source" and "Navigation".

At the bottom, the status bar indicates "7 COLUMNS, 999+ ROWS" and "Column profiling based on top 1000 rows". The Windows taskbar at the very bottom shows the system clock as 2:23 PM on 29-Jul-22.

Project Title :
Visualization of healthcare data in a specific geographical region
(Case study-Bihar)
K-Means Clustering

To train a clustering model we will execute Python script in
Power Query Editor (Power Query Editor → Transform →
Run python script).

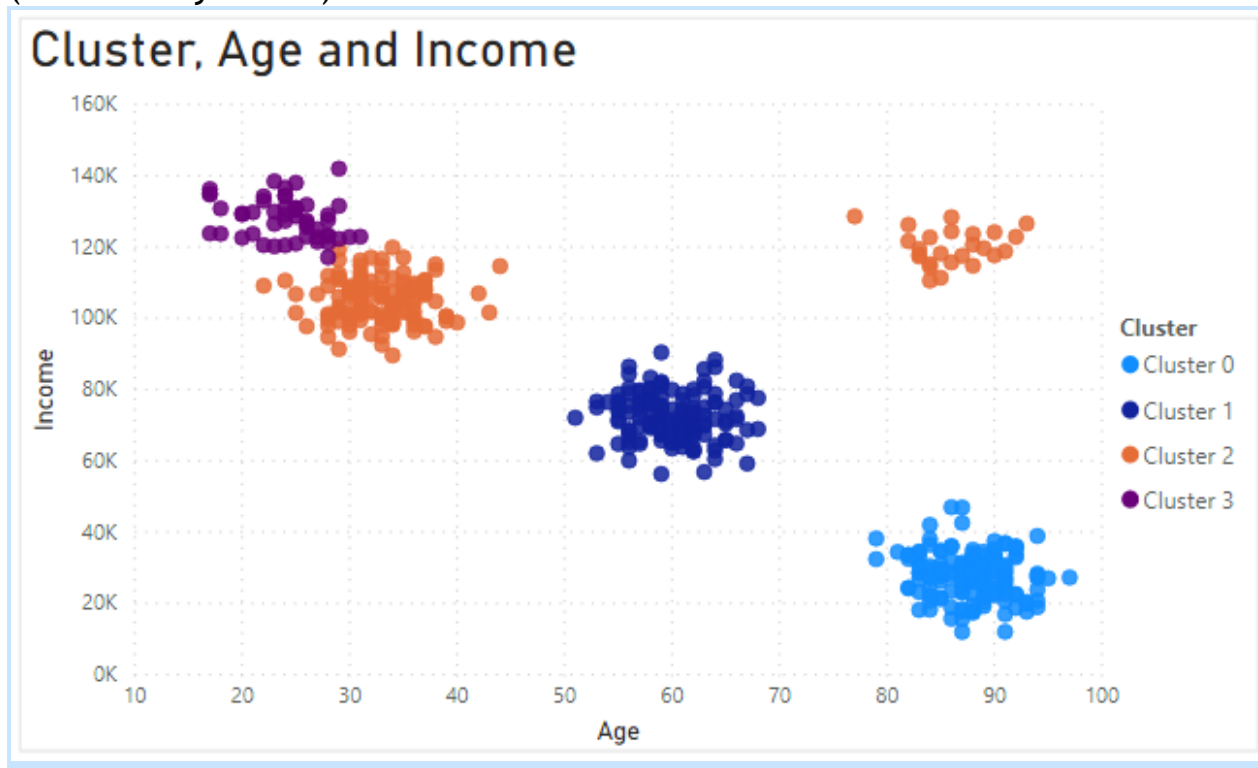
We run the following code as a python script:

```
from pycaret.clustering import *
```

```
dataset = get_clusters(data = dataset)
```

A new column '**Cluster**' containing label is attached to the
original table. Once you apply the query (Power Query Editor
→ Home → Close & Apply), Here is how you can visualize the
clusters in Power BI:

Project Title :
Visualization of healthcare data in a specific geographical region
(Case study-Bihar)



What's required to create good machine learning systems?

- Data preparation capabilities.
- Algorithms - basic and advanced.
- Automation and iterative processes.
- Scalability.
- Ensemble modeling.



Did you know?

- In machine learning, a target is called a label.
- In statistics, a target is called a dependent variable.
- A variable in statistics is called a feature in machine learning.
- A transformation in statistics is called feature creation in machine learning.

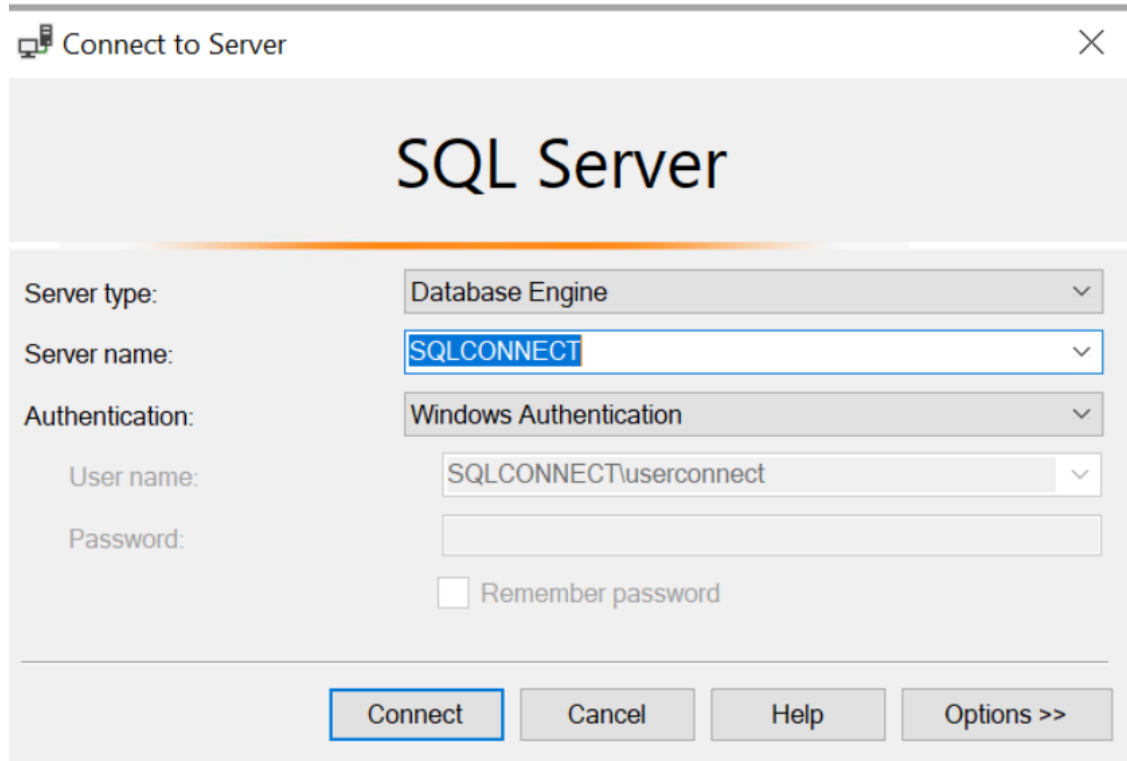
Project Title :

Visualization of healthcare data in a specific geographical region
(Case study-Bihar)

We used SQL Server Management Studio (SSMS) to connect to our SQL

Server instance. The procedure is explained below:

- Start SQL Server Management Studio. The first time you run SSMS, the Connect to Server window opens. If it doesn't open, you can open it manually by selecting Object Explorer > Connect > Database Engine.
- The Connect to Server dialog box appears. Enter the information accordingly.



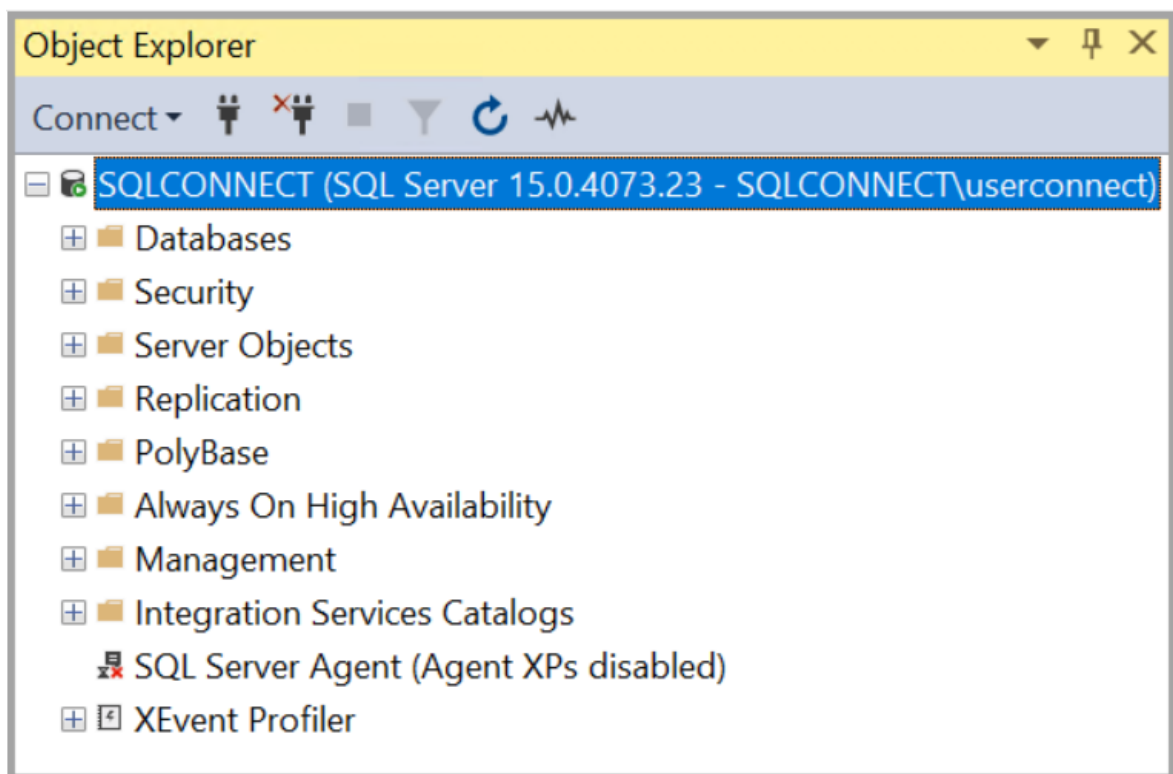
The screenshot shows the 'Connect to Server' dialog box in SQL Server Management Studio. The dialog box has a title bar with a close button (X) and the text 'Connect to Server'. The main area is titled 'SQL Server'. Below the title, there are several fields and options:

- Server type:** A dropdown menu showing 'Database Engine'.
- Server name:** A text box containing 'SQLCONNECT'.
- Authentication:** A dropdown menu showing 'Windows Authentication'.
- User name:** A text box containing 'SQLCONNECT\userconnect'.
- Password:** A text box (empty).
- Remember password:** A checkbox (unchecked).
- Buttons:** 'Connect', 'Cancel', 'Help', and 'Options >>'.

Project Title :

Visualization of healthcare data in a specific geographical region
(Case study-Bihar)

- After you've completed all the fields, select Connect.
- We can also modify additional connection options by selecting Options. Examples of connection options are the database we are connecting to, the connection timeout value, and the network protocol.
- To verify that our SQL Server connection succeeded, we expand and explore the objects within Object Explorer where the server name, the SQL Server version, and the username are displayed.



Project Title :
Visualization of healthcare data in a specific geographical region
(Case study-Bihar)

TASKS PERFORMED:

- Problem understanding and analysis
- Data collection, data prep/cleaning, and basic exploratory data analysis
- Model development and testing
- Model deployment, monitoring, and governance
- Communication of findings to business decision-makers

CHALLENGES FACED:

1. Inaccessible Data

Addressed by:

- Easily combining data from multiple, disparate sources into a virtual data layer
- Visually manipulating, cleaning, and transforming data to make it ready for analysis
- Using introspection and relationship discovery to understand and validate data relationships for model building

2. Dirty Data

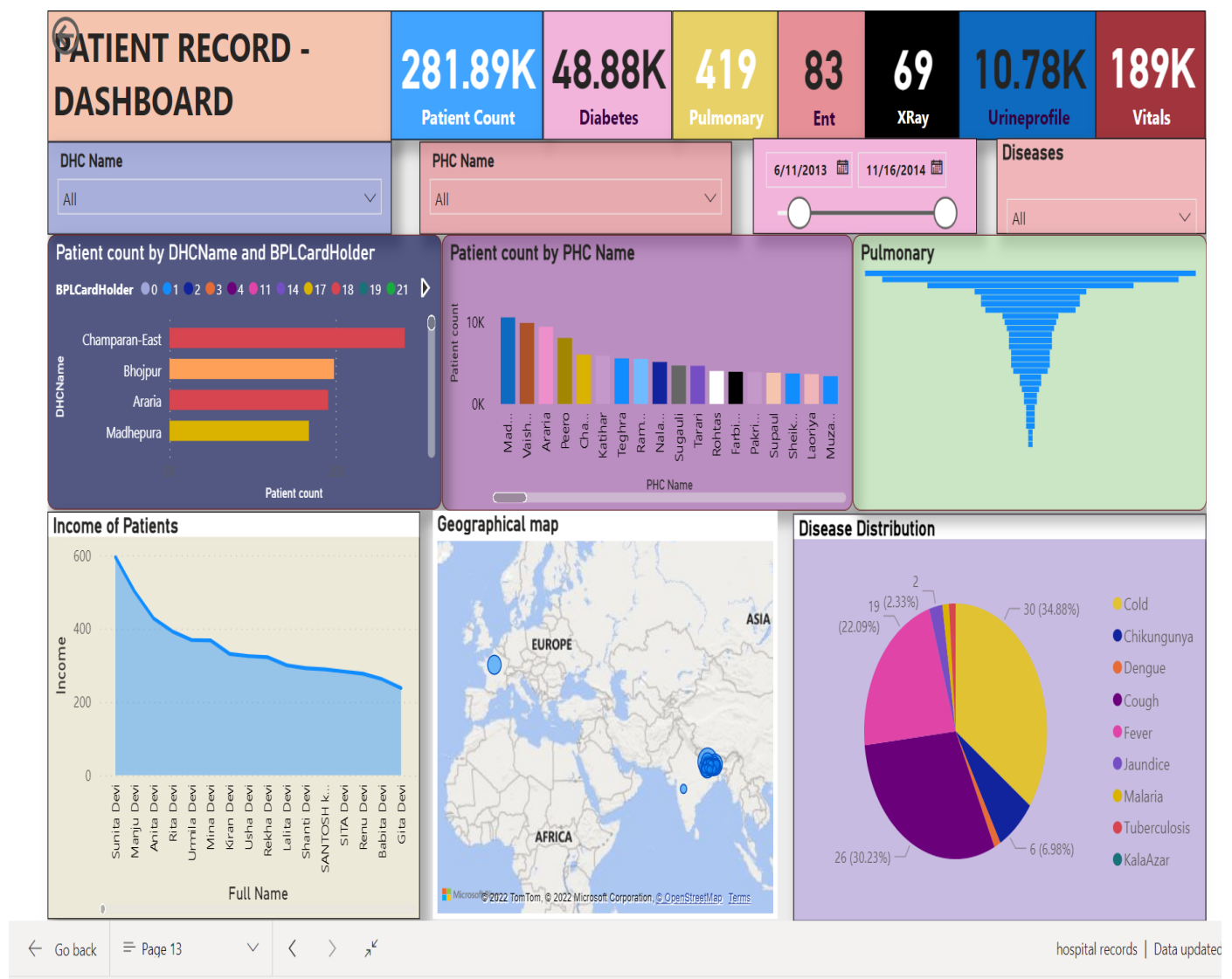
Addressed by:

- AI fueled visual wrangling to automatically suggest transformations, remove outliers, and clean data

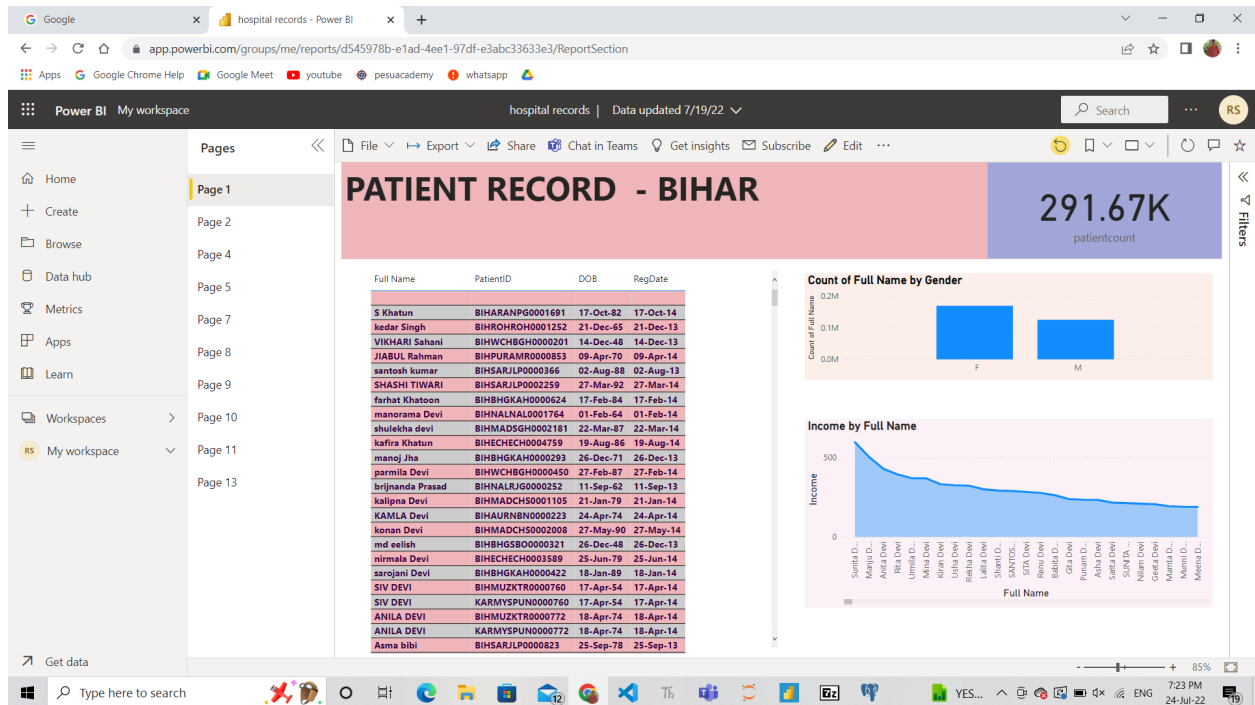
Project Title : Visualization of healthcare data in a specific geographical region (Case study-Bihar)

- Automated data health check to fill in missing values, remove unimportant variables, and prepare data for analytics
- Formatting and preparing data across disparate sources at scale

INTERACTIVE DASHBOARD OF HEALTHCARE DATA ON THE STATE OF BIHAR:



Project Title : Visualization of healthcare data in a specific geographical region (Case study-Bihar)



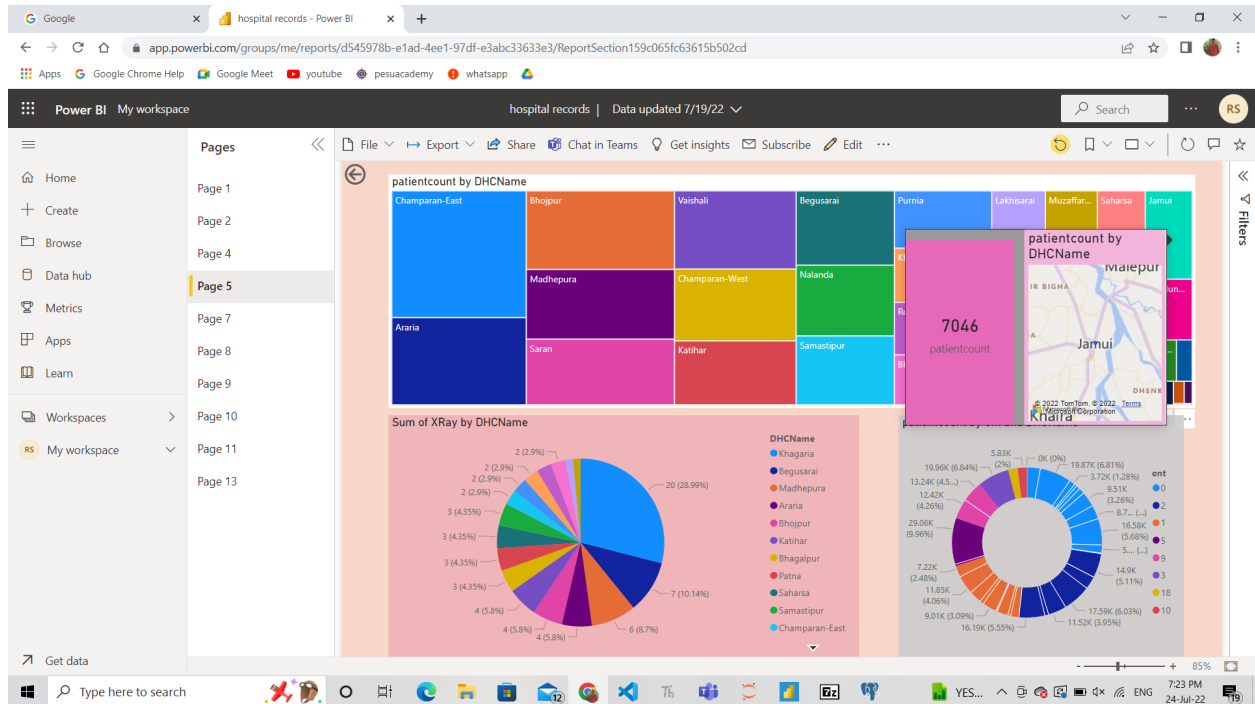
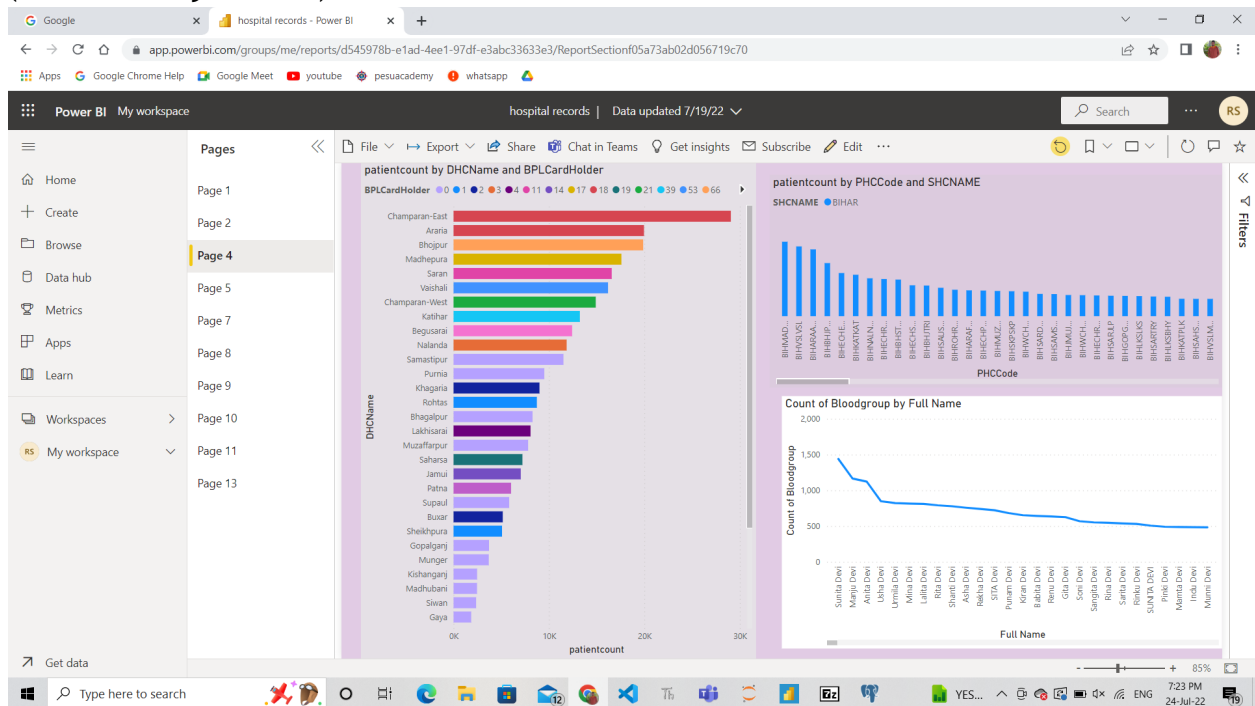
Google | hospital records - Power BI | +

app.powerbi.com/groups/me/reports/d545978b-e1ad-4ee1-97df-e3abc3363e3/ReportSection/c41a8eb000293b34de1

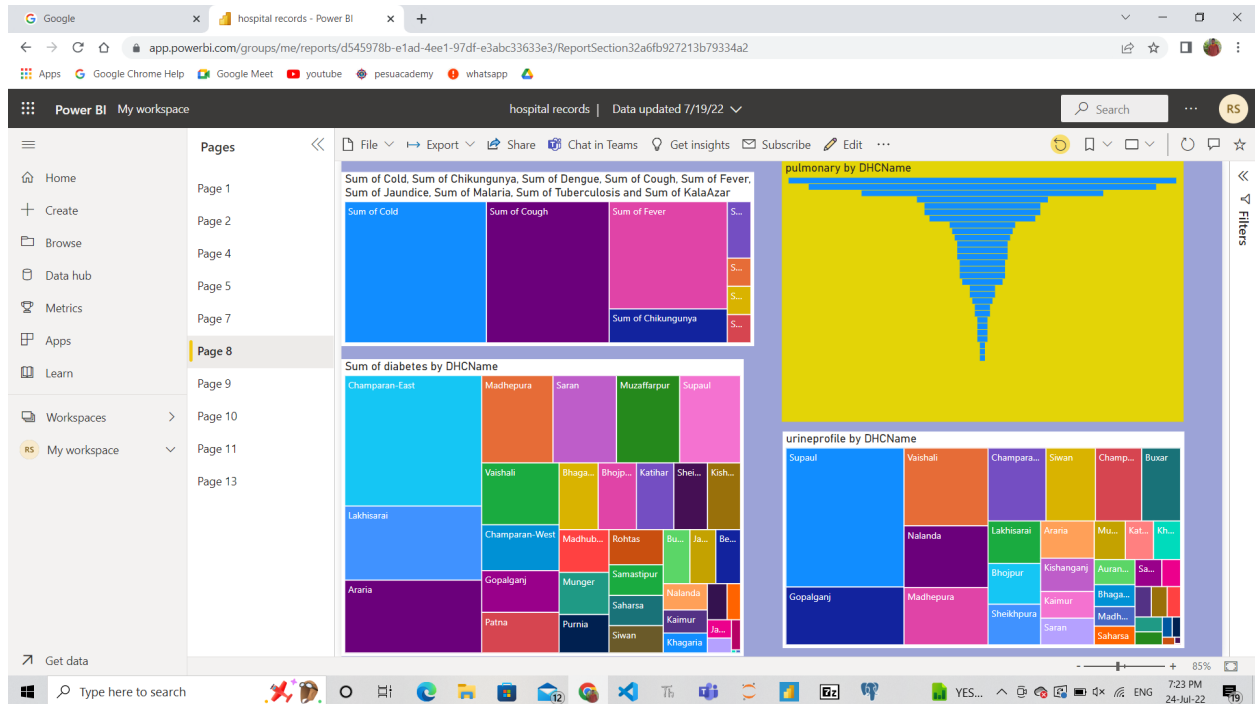
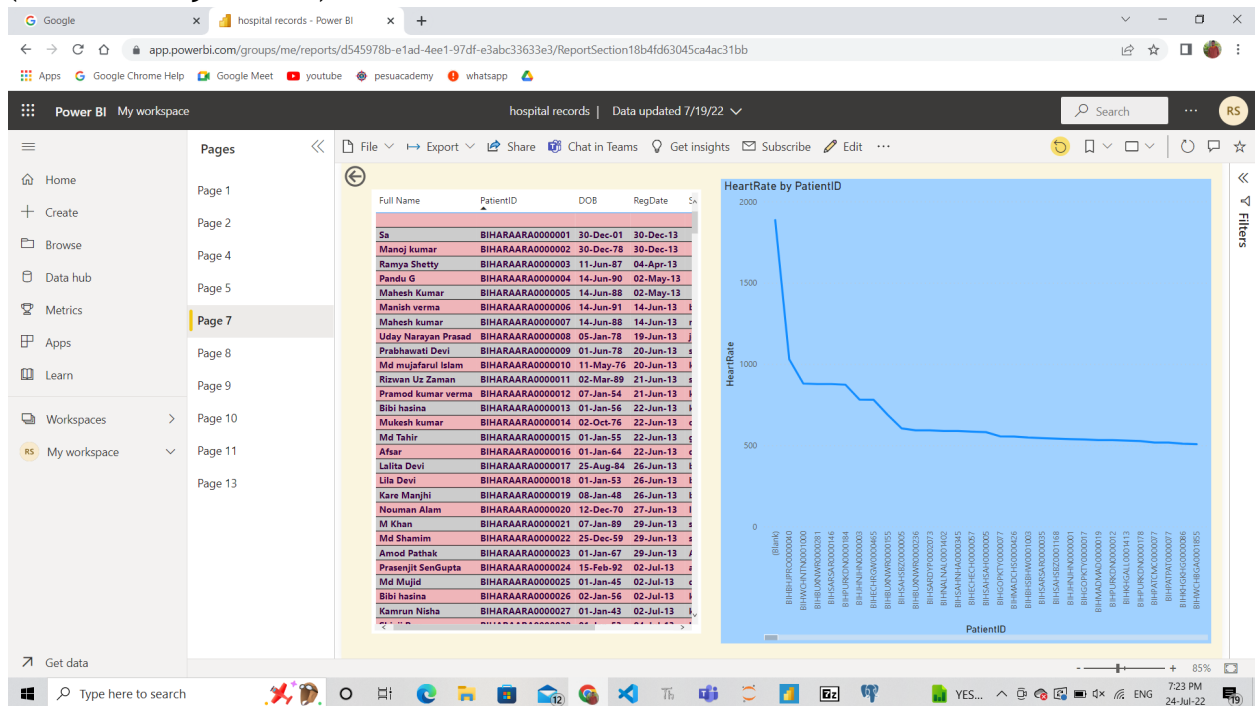
Power BI | My workspace | hospital records | Data updated 7/19/22

DHCName	PHCName	TotalCount	Male	Female	EarningthroughReferral	publicAddressedPatient	publicpatientinqueue	PrivateHealthProvider	AddressedPatient	patientin...
Araria	Araria	10200	2911	7289	1447	1385	62	24	17	
Madhepura	Madhepura	11363	5035	6328	112	91	21	0	0	
Champanan-East	Champanan-East	15681	3122	3459	46	16	30	58	57	
Vaishali	Vaishali	10646	5414	5232	77	34	43	0	0	
Bhojpur	Peero	8566	3648	4917	0	0	0	0	0	
Rohitas	Rohitas	4033	1387	2646	14	10	2	1	0	
Nalanda	Nalanda	5778	2105	3673	342	218	124	17	0	
Champanan-West	Laoriya	3737	1099	2638	0	0	0	0	0	
Sheikhpura	Sheikhpura	3004	1685	2119	13	0	13	0	0	
Katihar	Katihar	6321	3496	2825	997	973	24	0	0	
Muzaffarpur	Muzaffarpur	3945	1882	1963	266	251	15	63	62	
Jamul	Jamul	3225	611	2614	192	119	73	64	42	
Champanan-East	Ramgarhawa	5670	1760	3910	0	0	0	0	0	
Saran	Jalalpur	3110	1274	1836	1	0	1	0	0	
Gopiganj	Gopiganj	3094	1835	1259	78	40	38	6	0	
Supaul	Supaul	4332	2451	1881	15	12	3	11	11	
Saran	Tairalya	2995	912	2083	4	4	0	0	0	
Lakhisaral	Lakhisaral	3062	1639	1423	0	0	0	13	13	
Champanan-East	Raxaul	3162	1407	1755	0	0	0	0	0	
Lakhisaral	Barahia	2963	1448	1515	0	0	0	0	0	
Champanan-East	Sugauli	4709	1891	2818	0	0	0	0	0	
Champanan-West	Champanan-West	3201	1232	1969	3	3	0	0	0	
Jamul	Khaira	2447	1015	1432	0	0	0	0	0	
Khagaria	Khagaria	2498	1369	1129	315	256	59	1	0	
Semastipur	Semastipur	3370	1729	1641	15	15	0	0	0	
Saharsa	Saharsa	2638	1559	1079	0	0	0	0	0	
Purnia	Banmankhi	2336	1262	1074	0	0	0	0	0	
Patna	Naubatpur	2604	633	1971	0	0	0	0	0	
Saran	Parsa	2172	954	1218	0	0	0	0	0	
Madhepura	Chausa	2375	948	1427	5	0	5	0	0	
Saran	Garkha	1976	697	1279	0	0	0	0	0	
Total		291770	123864	167906	5769	4895	869	503	386	

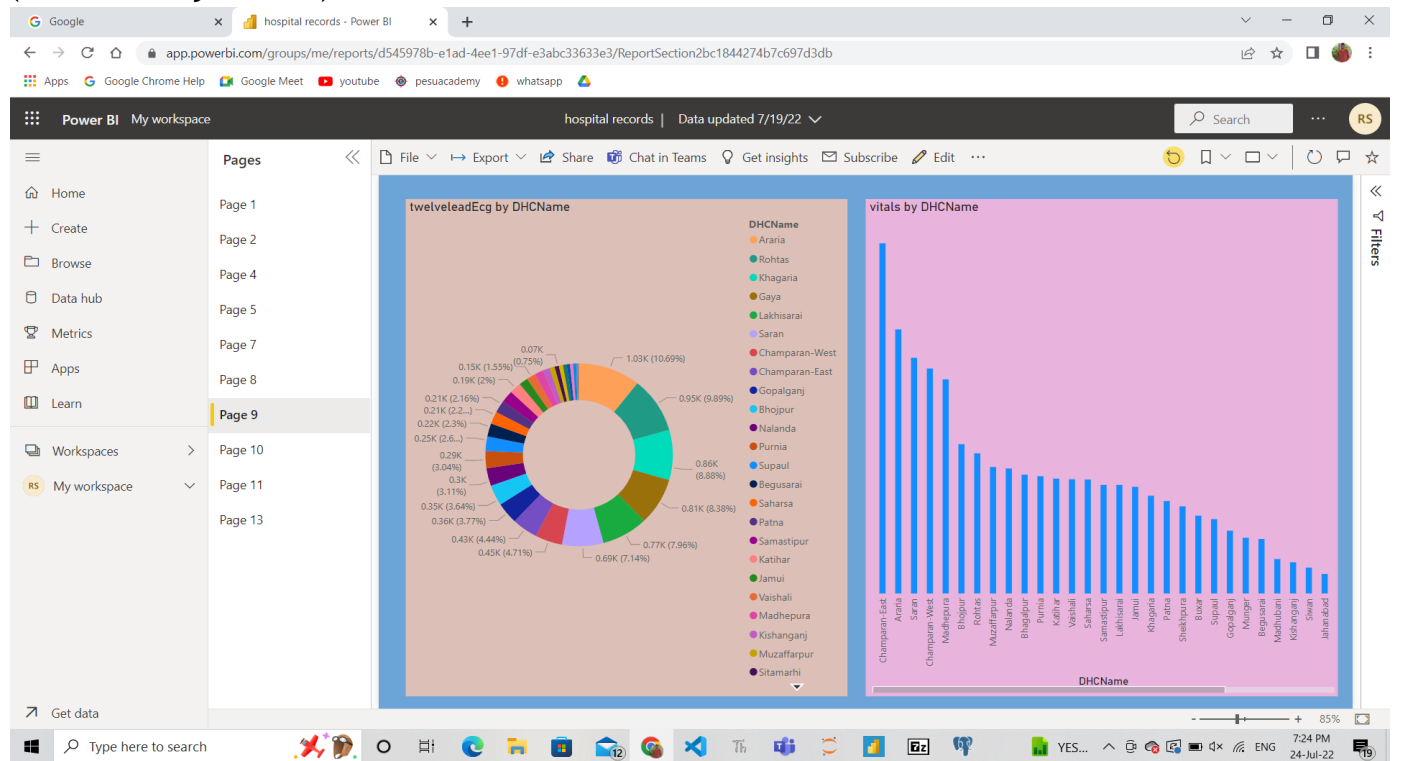
Project Title : Visualization of healthcare data in a specific geographical region (Case study-Bihar)



Project Title : Visualization of healthcare data in a specific geographical region (Case study-Bihar)



Project Title : Visualization of healthcare data in a specific geographical region (Case study-Bihar)



Google | hospital records - Power BI

app.powerbi.com/groups/me/reports/d545978b-e1ad-4ee1-97df-e3abc33633e3/ReportSection3050a9e34da3b43201c0

Power BI | My workspace | hospital records | Data updated 7/19/22

Pages: Page 1, Page 2, Page 4, Page 5, Page 7, Page 8, Page 9, Page 10, Page 11, Page 13

PatientID	DATETIME	ECGHour	ECGMin	ECGSec	HeartRate	ImageId	ProductID
BIHARAARA0001879	1/4/2014 11:21:52 AM	0	0	25	84	0001	IMGECG
BIHARAARA0001882	1/4/2014 3:55:27 PM	0	0	25	100	0001	IMGECG
BIHARAARA0001886	1/4/2014 4:33:53 PM	0	0	25	0	0001	IMGECG
BIHARAARA0001911	1/6/2014 4:02:09 PM	0	0	26	69	0001	IMGECG
BIHARAARA0002003	1/10/2014 9:57:14 AM	0	0	25	69	0001	IMGECG
BIHARAARA0002018	1/10/2014 4:05:44 PM	0	0	25	66	0001	IMGECG
BIHARAARA0002032	1/11/2014 3:57:56 PM	0	0	25	0	0001	IMGECG
BIHARAARA0002056	1/13/2014 11:48:42 AM	0	0	25	69	0001	IMGECG
BIHARAARA0002059	1/13/2014 12:10:34 PM	0	0	25	73	0001	IMGECG
BIHARAARA0002081	1/16/2014 9:26:29 AM	0	0	25	82	0001	IMGECG
BIHARAARA0002086	1/16/2014 11:52:37 AM	0	0	24	0	0001	IMGECG
BIHARAARA0002118	1/16/2014 12:32:40 PM	0	0	25	100	0001	IMGECG
BIHARAARA0002207	1/18/2014 11:53:36 AM	0	0	25	60	0001	IMGECG
BIHARAARA0002213	1/18/2014 5:04:53 PM	0	0	25	76	0001	IMGECG
BIHARAARA0002248	1/21/2014 9:27:16 AM	0	0	25	100	0001	IMGECG
BIHARAARA0002253	1/21/2014 10:16:27 AM	0	0	25	60	0001	IMGECG
BIHARAARA0002304	1/23/2014 10:57:57 AM	0	0	25	90	0001	IMGECG
BIHARAARA0002328	1/23/2014 11:10:29 AM	0	0	25	71	0001	IMGECG
BIHARAARA0002353	1/23/2014 12:35:07 PM	0	0	25	89	0001	IMGECG
BIHARAARA0002357	1/23/2014 4:09:12 PM	0	0	25	109	0001	IMGECG
BIHARAARA0002358	1/23/2014 5:36:21 PM	0	0	25	81	0001	IMGECG
BIHARAARA0002500	1/28/2014 4:59:37 PM	0	0	25	70	0001	IMGECG
BIHARAARA0002563	1/30/2014 11:17:37 AM	0	0	25	89	0001	IMGECG
BIHARAARA0002588	1/31/2014 11:18:19 AM	0	0	25	69	0001	IMGECG
BIHARAARA0002599	1/31/2014 12:16:42 PM	0	0	25	120	0001	IMGECG
BIHARAARA0007622	2/1/2014 11:41:11 AM	0	0	25	87	0001	IMGECG

Mail

Mail has new messages

24°C

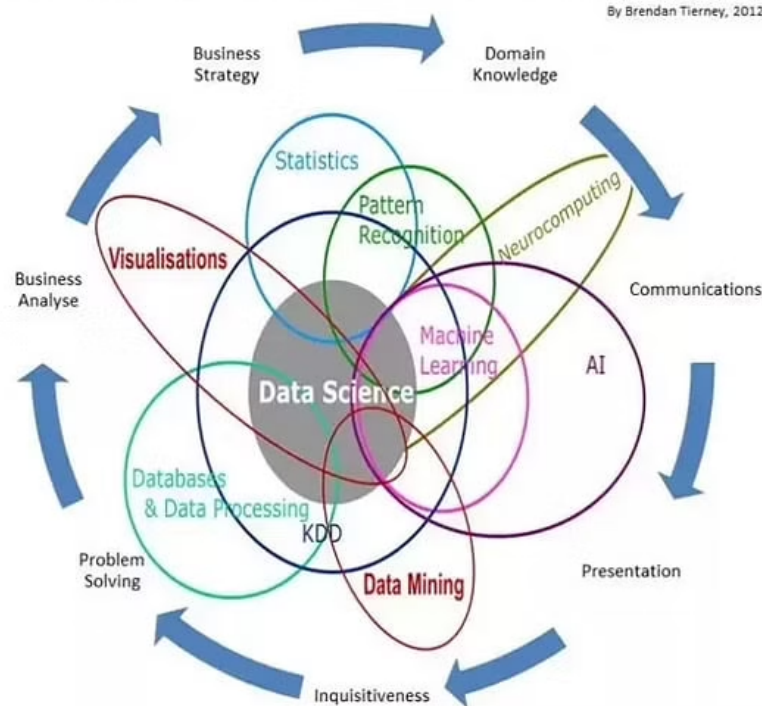
7:24 PM

24-Jul-22

Project Title :
Visualization of healthcare data in a specific geographical region
(Case study-Bihar)

CONCLUSION:

Data Science Is Multidisciplinary



The data science process is simple to understand and involves the following steps:

1. Understand the business problem
2. Gather and integrate the raw data
3. Explore, transform, clean, and prepare the data
4. Create and select models based on the data
5. Test, tune, and deploy the models
6. Monitor, test, refresh, and govern the models

Project Title :
Visualization of healthcare data in a specific geographical region
(Case study-Bihar)

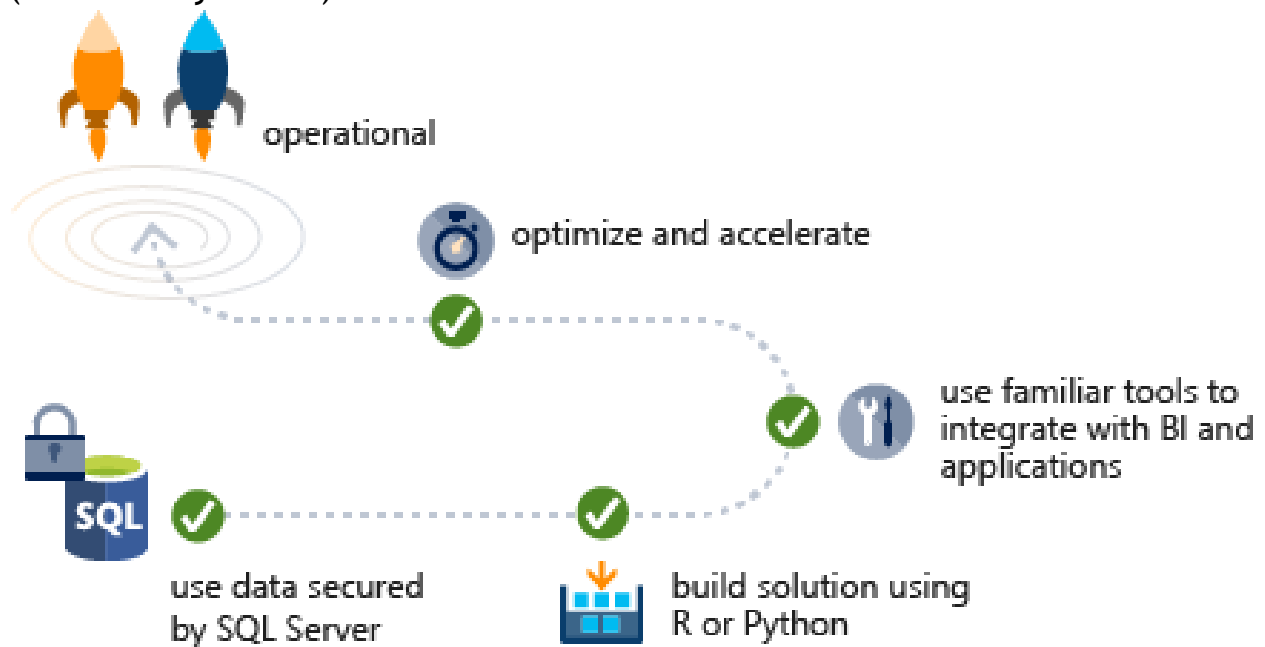
Understand the Machine Learning Process

- Data acquisition and integration
- Data exploration, preparation, and cleansing
- Data preprocessing, transformation, and feature generation
- Model development and selection
- Model testing and tuning
- Model deployment

Understand the Model Operations and Governance Process

- Model repository, documentation, and version control
- Model scoring, API framework, and container strategy
- Model execution environment
- Model deployment, integration, and results
- Model monitoring, testing, and refresh

Project Title :
Visualization of healthcare data in a specific geographical region
(Case study-Bihar)



KEY BENEFITS OF MACHINE LEARNING SERVICES IN SQL SERVER:

- Avoids data movement and data risk by bringing analytics to the data
- Integrates the best features of open source with Microsoft's enterprise capabilities (revoscale and microsoftml libraries).
- Easy to deploy and consume predictive models
- Integration with existing workflows
- Unified governance across analytics and storage

There are two platforms introduced by Microsoft to integrate with open source R/Python languages with business applications:

Project Title :

Visualization of healthcare data in a specific geographical region
(Case study-Bihar)

- SQL Server Machine Learning Services (In-Database) supports both R and Python for in-database analytics
- Microsoft Machine Learning Server supports R and Python deployments on Windows, Linux, HDInsight Spark and Hadoop clusters.

DATA STREAMING IN POWER BI:

- REAL TIME DASHBOARDS:Real-time dashboards Power BI lets you easily display and analyze your realtime data, empowering your organization to gain instant insights from time-sensitive information. Monitor social media campaigns as they go viral. Display streaming video on your dashboards. Bring your dashboards to life with IoT sensor readouts. The rich functionality of Power BI, combined with the velocity of real-time data, will transform the way you do business
- Setup of real-time streaming data sets With Power BI real-time streaming, you can stream data and update dashboards in real time. Any visual or dashboard that can be created in Power BI can also be created to display and update real-time data and visuals.