# NAAN MUDHALVAN PROJECT BASED LEARNING

## MEENAKSHI SUNDARARAJAN ENGINEERING COLLEGE

Kodambakkam, Chennai-600024.

### DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

**TOPIC: Estimation Of Business Project**

TEAM ID: NM2023TMID07219

FACULTY MENTOR: YAMUNA S

INDUSTRY MENTOR: SHIVAM SHIVHARE

Project submitted by,

HARINIESWARI V(311520104018),
SRIMATHI T(311520104045),
VAISHNAVI R(311520104055)

# ABSTRACT

The objective of this data analytics project is to unveil nuanced insights into the financial dynamics of educational institutions, with a specific focus on operating expenses. The project employs a comprehensive approach, dissecting various expenditure categories from 'Annual Payroll' to 'Other Operating Expenses,' providing stakeholders with actionable insights to optimize resource utilization and foster financial sustainability within the education sector.The dataset encompasses a diverse array of expenditure categories, offering a detailed examination of total expenses and their percentage distribution. Critical components such as 'Detailed Employer Costs for Fringe Benefits,' 'Purchased Communication Services,' and 'Depreciation and Amortization Charges' are thoroughly analyzed to provide a holistic understanding of financial patterns within the education sector.Utilizing DB2 metrics and responsive data handling strategies, the project ensures timely access to financial data. Time-based, institutional, and category-specific filters enhance the granularity of the analysis, empowering stakeholders to customize their exploration of operating expenses.Real-time Data Insights: Implementing mechanisms for real-time data updates and rendering, ensuring stakeholders access the most current financial information.Query Performance Optimization: Utilizing DB2 metrics to optimize query execution, enhance data retrieval speed, and improve overall system responsiveness.Interactive Data Exploration: Incorporating user-friendly filters for time, institution, and expense categories, allowing stakeholders to interactively explore and analyze operating expenses.Customized Insights: Providing flexibility through custom filters, enabling stakeholders to tailor analyses based on specific criteria relevant to their goals.Cross-Dimensional Analysis: Implementing cross-filtering capabilities to reveal relationships within the data, enhancing the depth of insights derived from the dataset.The data analytics report aims to provide actionable recommendations for stakeholders in the education sector. By understanding the financial dynamics and discerning patterns within each expense category, institutions can make informed decisions, optimize resource utilization, and foster long-term financial sustainability.This project employs advanced data analytics techniques, DB2 metrics, and customizable filters to present a comprehensive analysis of operating expenses in the education sector. The outcomes are intended to guide stakeholders in optimizing financial strategies, ultimately contributing to the overall financial health and sustainability of educational institutions.

# Project Report Format

1. **INTRODUCTION**
   - 1.1 Project Overview
   - 1.2 Purpose
2. **LITERATURE SURVEY**
   - 2.1 Existing problem
   - 2.2 References
   - 2.3 Problem Statement Definition
3. **IDEATION & PROPOSED SOLUTION**
   - 3.1 Empathy Map Canvas
   - 3.2 Ideation & Brainstorming
   - 3.3 Proposed Solution
   - 3.4 Problem Solution fit
4. **REQUIREMENT ANALYSIS**
   - 4.1 Functional requirement
   - 4.2 Non-Functional requirements
5. **PROJECT DESIGN**
   - 5.1 Data Flow Diagrams
   - 5.2 Solution & Technical Architecture
   - 5.3 User Stories
6. **PROJECT PLANNING & SCHEDULING**
   - 6.1 Sprint Planning & Estimation
   - 6.2 Sprint Delivery Schedule
   - 6.3 Reports from JIRA
7. **CODING & SOLUTION (Explain the features added in the project along with code)**
   - 7.1 Feature 1
   - 7.2 Feature 2
   - 7.3 Database Schema (if Applicable)
8. **TESTING**
   - 8.1 Test Cases
   - 8.2 User Acceptance Testing
9. **RESULTS**
   - 9.1 Performance Metrics
10. **ADVANTAGES & DISADVANTAGES**
11. **CONCLUSION**
12. **FUTURE SCOPE**
13. **APPENDIX**
    - Source Code
    - GitHub & Project Demo Link

# INTRODUCTION

In the ever-evolving landscape of education, understanding and optimizing financial resources are integral to the sustainability and growth of educational institutions. This data analytics project embarks on a journey to unravel the intricate financial dynamics within the education sector, with a laser focus on operating expenses. By employing cutting-edge techniques and leveraging the power of DB2 metrics, the project aims to provide stakeholders with a detailed, data-driven narrative that goes beyond mere numbers.

Educational institutions, ranging from schools to universities, operate within a complex web of financial intricacies. The allocation of resources, particularly in the realm of operating expenses, plays a pivotal role in shaping the quality of education, the well-being of faculty and staff, and the overall effectiveness of institutional operations. In this context, a granular analysis becomes imperative to unearth hidden patterns, discern trends, and extract actionable insights.

Project Objectives:

The primary objective of this data analytics endeavor is to deliver a comprehensive understanding of the financial fabric that weaves through educational institutions. Specifically, the project aims to:

Examine Operating Expenses: Dive deep into various operating expense categories, from payroll to fringe benefits, communication services, and beyond, to illuminate the financial landscape comprehensively.

Utilize DB2 Metrics: Harness the power of DB2 metrics to ensure not only real-time data responsiveness but also optimized query performance, enabling stakeholders to interact with data seamlessly.

Implement Customizable Filters: Introduce a dynamic filtering system that empowers stakeholders to tailor their analyses, exploring financial data based on time, institutions, expense categories, and other parameters.

Derive Actionable Insights: Uncover actionable insights that guide stakeholders in resource optimization, decision-making, and the pursuit of financial sustainability within the education sector.

Significance of the Project:

This project goes beyond traditional financial reports by bringing forth a responsive and interactive approach to data analytics. In a sector where every dollar spent can influence the quality of education, the insights derived from this project have the potential to reshape financial strategies, enhance resource utilization, and contribute to the resilience of educational institutions.

The forthcoming sections will delve into the nuances of operating expenses, employing a data-centric lens. From real-time data updates to the utilization of customizable filters, the report will unfold a detailed narrative designed to empower stakeholders with actionable insights, fostering financial health and sustainability in the education sector.

## PROJECT OVERVIEW

In presenting a data analytic report focusing on the columns and subcolumns related to operating expenses in the education sector, the objective is to unveil intricate insights into the financial dynamics of educational institutions. The dataset encapsulates a diverse array of expenditure categories, ranging from 'Annual Payroll' to 'Purchased Professional and Technical Services.' This thorough examination encompasses not only the total expenses but also their percentage distribution, providing a nuanced understanding of resource allocation. The breakdown of 'Detailed Employer Costs for Fringe Benefits' delves into critical components like health insurance and pension plans, shedding light on the broader scope of employee benefits. Moreover, the analysis explores costs related to 'Purchased Communication Services,' 'Depreciation and Amortization Charges,' and 'Other Operating Expenses.' This comprehensive approach not only dissects the financial fabric of education but also aims to discern patterns and trends within each expense category. The data analytics report endeavors to present actionable insights, guiding stakeholders in optimizing resource utilization and fostering financial sustainability within the education sector.

## PROJECT FLOW

- Define Problem / Problem Understanding
- Specify the business problem
- Business requirements
- Literature Survey
- Social or Business Impact.
- Data Collection & Extraction from Database
- Collect the dataset,
- Storing Data in DB2
- Perform SQL Operations
- Connect DB2 with Cognos
- Data Preparation
- Prepare the Data for Visualization
- Data Visualizations
- No of Unique Visualizations
- Dashboard
- Responsive and Design of Dashboard
- Story
- No of Scenes of Story
- Report
- No of Visualization with detail information
- Performance Testing
- Amount of Data Rendered to DB2
- Utilization of Data Filters
- No of Calculation Fields
- No of Visualizations/ Graphs
- Web Integration
- Dashboard, Report and Story embed with UI With Flask
- Project Demonstration & Documentation
- Record explanation Video for project end to end solution
- Project Documentation-Step by step project development procedure

**PURPOSE**

The purpose of this data analytics project is multifaceted, with a strategic focus on illuminating the financial dynamics within the education sector. The overarching goals revolve around providing stakeholders with actionable insights that contribute to the financial sustainability and optimization of resources within educational institutions.

## LITERATURE SURVEY

A literature survey for a project on analyzing operating expenses in the education sector would involve reviewing existing research, studies, and scholarly works that provide insights into financial dynamics, data analytics in education, and related areas. Below is a structured literature survey that covers key themes relevant to your project:

Financial Dynamics in Education:

Explore literature that delves into the financial dynamics of educational institutions, emphasizing studies that focus on budgeting, financial management, and resource allocation in the education sector.

Operating Expenses in Education:

Review research that specifically addresses operating expenses in educational settings. Look for studies that discuss trends, challenges, and best practices in managing and optimizing operating budgets.

Data Analytics in Education:

Investigate literature related to the application of data analytics in the education sector. This should include studies that highlight the use of data-driven approaches to enhance decision-making, resource allocation, and financial sustainability.

Database Management and Metrics:

Explore scholarly works that discuss best practices in database management, especially in the context of financial data in educational institutions. Look for insights on optimizing database performance, utilizing metrics, and ensuring data responsiveness.

Customizable Filters in Data Analytics:

Review literature that discusses the implementation and impact of customizable filters in data analytics. This could include studies on user interface design, interactive data exploration, and the effectiveness of filtering mechanisms in enhancing data analysis.

Real-Time Data Analytics:

Investigate how real-time data analytics is applied in various sectors, with a focus on studies that highlight its relevance and benefits in the education sector. Consider literature that discusses real-time updates, query optimization, and data responsiveness.

Financial Sustainability in Education:

Look for literature that addresses the concept of financial sustainability in education. Explore studies that propose frameworks, models, or strategies for achieving financial sustainability in educational institutions.

Decision-Making in Education:

Review research on decision-making processes within educational institutions, especially those that emphasize the role of data-driven insights in shaping financial decisions. Consider how analytics

can inform strategic planning.

# EXISTING PROBLEM AND REFERENCES

## Problem Statement Definition

In the initial phase of our "Business Estimation" project, Milestone 1 is dedicated to defining the problem and cultivating a profound understanding of the challenges associated with estimating business expenses. The problem statement encapsulates the need to bridge the gap between current estimations of operating expenses and the desired state of accurate and informed financial projections. This involves identifying discrepancies and inefficiencies within the current estimation processes and elucidating the significance of addressing these issues. The impact of this problem on businesses is emphasized, highlighting the importance of precise financial estimations for strategic decision-making and overall organizational efficiency.

# IDEATION & PROPOSED SOLUTION

## Empathy Map Canvas


Empathy Map Canvas is a useful tool for understanding and empathizing with the needs, thoughts, and emotions of a specific target audience. An empathy map for CrimeVision, an advanced crime classification system that utilizes deep learning:

Creating an empathy map involves understanding and documenting the thoughts, feelings, actions, and pain points of the various stakeholders involved in your project. In the context of your data analytics project on operating expenses in the education sector, here's an empathy map that focuses on key stakeholders:

Stakeholders:

Administrators/Decision-Makers:

Thoughts: Concerned about optimizing financial resources for the institution.
Feelings: Pressure to make informed decisions for the financial sustainability of the institution.
Actions: Actively seeking data-driven insights for budgeting and resource allocation.

Pain Points: Limited visibility into nuanced financial patterns and potential inefficiencies.
Financial Planners/Analysts:

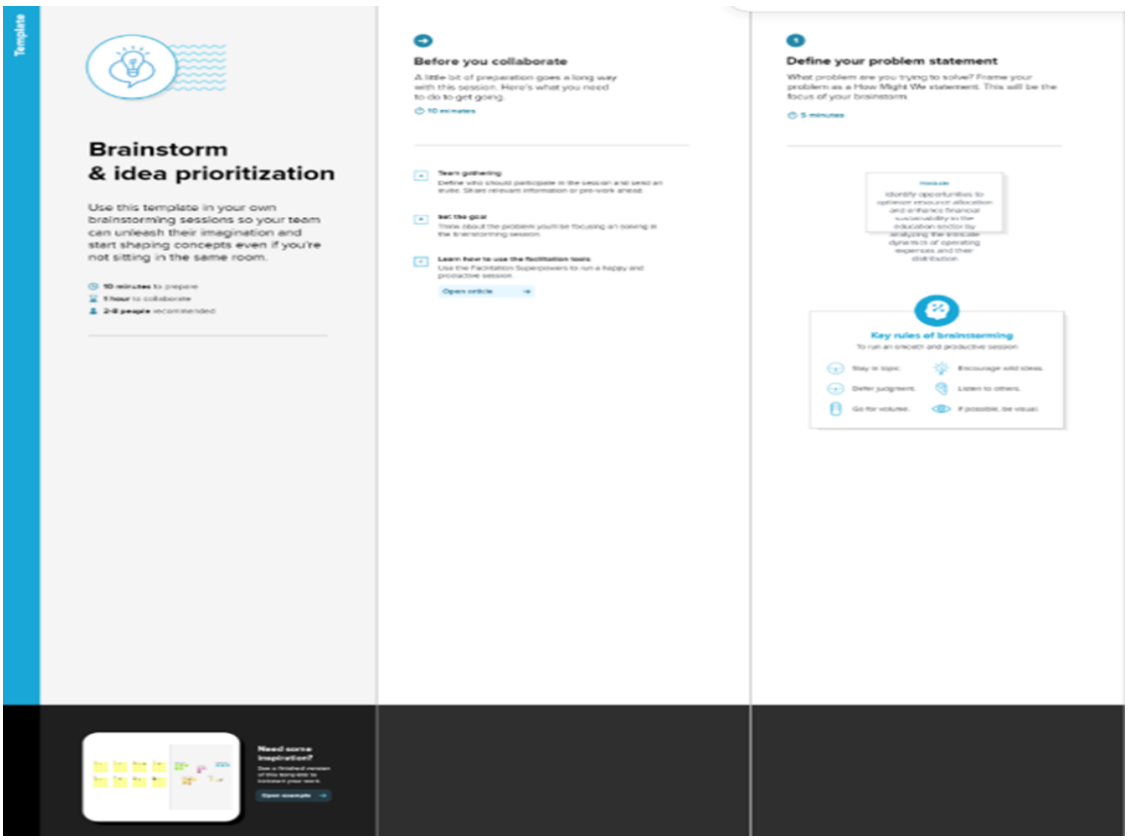
Thoughts: Interested in uncovering actionable insights within financial data.
Feelings: Enthusiastic about using analytics to enhance financial planning.
Actions: Actively analyzing data, looking for trends, and seeking optimization opportunities.
Pain Points: Challenges in accessing real-time data and comprehensive analytics tools.
Educational Institution Staff:

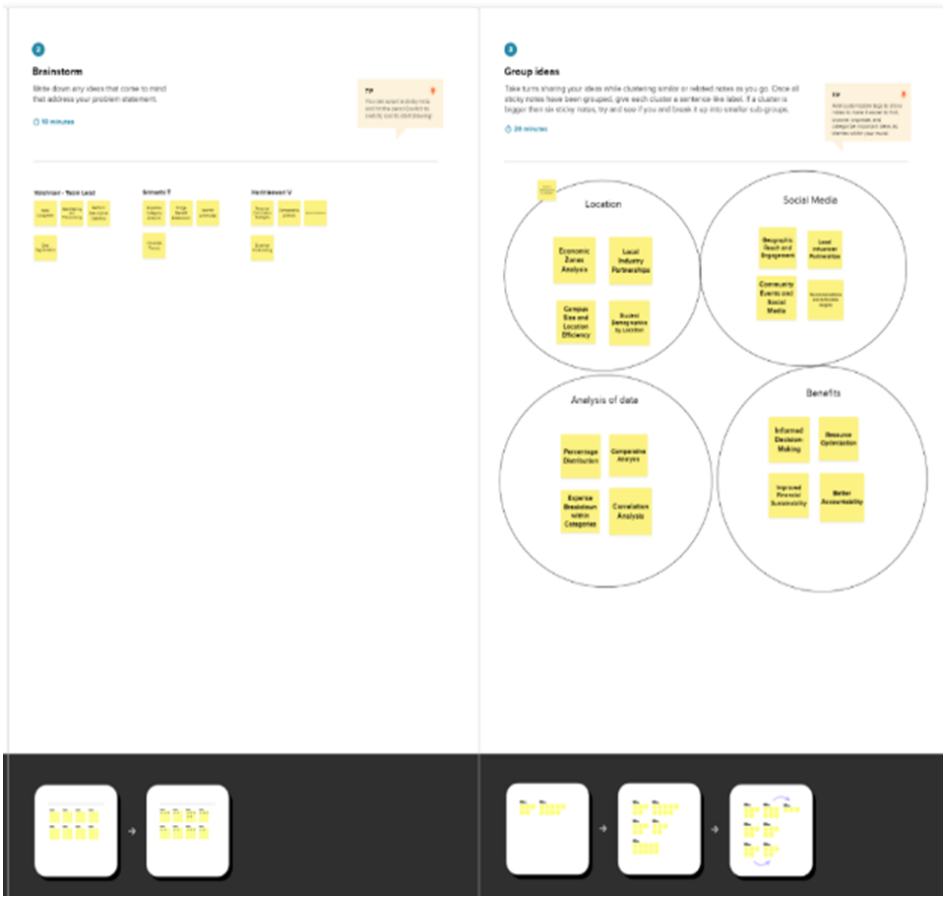
Thoughts: Interested in understanding how budget decisions impact day-to-day operations.
Feelings: Concerned about potential changes in resource allocation affecting their roles.
Actions: Providing input on departmental needs and resource requirements.
Pain Points: Uncertainty about how budget decisions are made and potential impacts on work.
Regulatory Compliance Officers:

Thoughts: Interested in ensuring financial practices align with regulatory standards.
Feelings: Focused on maintaining compliance and minimizing financial risks.
Actions: Reviewing financial data to ensure adherence to regulations.
Pain Points: Lack of clarity on specific financial practices and potential compliance issues.
Data Analysts/IT Professionals:

Thoughts: Focused on optimizing database performance and ensuring data responsiveness.
Feelings: Motivated to implement cutting-edge data analytics techniques.
Actions: Implementing and monitoring database configurations, optimizing queries.
Pain Points: Technical challenges related to data responsiveness and optimization.

## Ideation & Brainstorming:

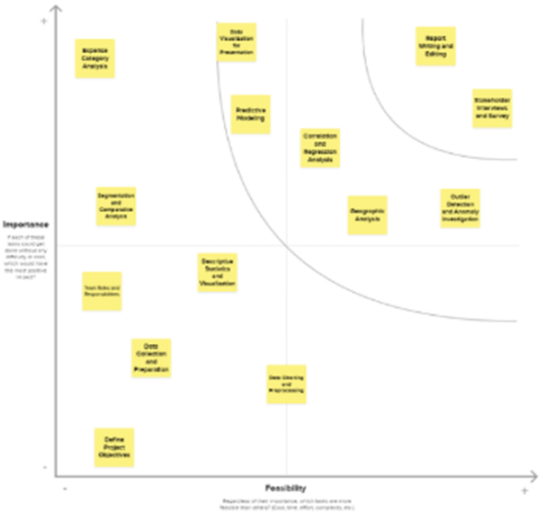



2
Brainstorm
Write down any ideas that come to mind that address your problem statement.
10 minutes
TIP
3
Group ideas
Take turns sharing your ideas while clustering similar or related notes as you go. Once all sticky notes have been grouped, give each cluster a sentence-like label. If a cluster is bigger than six sticky notes, try and see if you and break it up into smaller sub-groups.
20 minutes
TIP
Location
Economic Zones Analysis
Local Industry Partnerships
Campus Size and Location Efficiency
Student Demographics by Location
Social Media
Geographic Reach and Engagement
Local Influencer Partnerships
Community Events and Social Media
Analysis of data
Percentage Distribution
Comparative Analysis
Expense Breakdown within Categories
Correlation Analysis
Benefits
Informed Decision-Making
Resource Optimization
Improved Financial Sustainability
Better Accountability

## Prioritize

Your team should all be on the same page about what's important moving forward. Place your ideas on this grid to determine which ideas are important and which are feasible.

20 minutes




## After you collaborate

You can export the mural as an image or pdf to share with members of your company who might find it helpful.

### Quick add-ons

**Share the mural**
**Share a view link** to the mural with stakeholders to keep them in the loop about the outcomes of the session.

**Export the mural**
Export a copy of the mural as a PNG or PDF to attach to emails, include in slides, or save in your drive.

### Keep moving forward

**Strategy blueprint**
Define the components of a new idea or strategy.
Open the template →

**Customer experience journey map**
Understand customer needs, motivations, and obstacles for an experience.
Open the template →

**Strengths, weaknesses, opportunities & threats**
Identify strengths, weaknesses, opportunities, and threats (SWOT) to develop a plan.
Open the template →

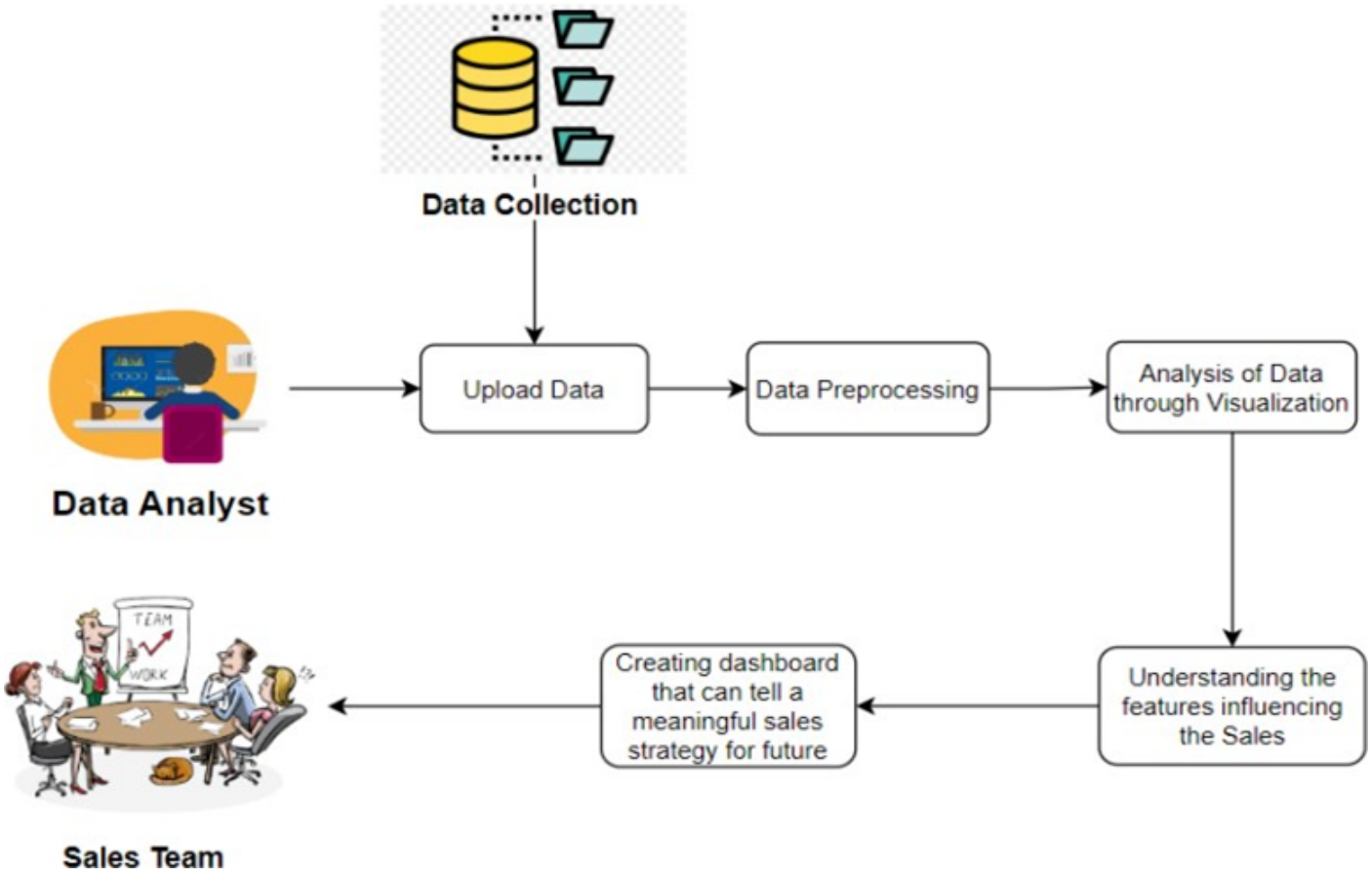
Share template feedback

## PROPOSED SOLUTION

Solution Architecture:

Solution architecture is a complex process – with many sub-processes – that bridges the gap between business problems and technology solutions. Its goals are to:

- Find the best tech solution to solve existing business problems.
- Describe the structure, characteristics, behavior, and other aspects of the software to project stakeholders.
- Define features, development phases, and solution requirements.
- Provide specifications according to which the solution is defined, managed, and delivered.




The social and business impact of accurate business estimation is multifaceted. On a social level, it contributes to increased transparency and accountability in financial reporting, fostering trust among stakeholders. Moreover, precise business estimation positively influences organizational competitiveness and efficiency. It aids businesses in making informed strategic decisions, optimizing resource allocation, and ultimately enhancing their overall financial health. The project's impact extends beyond mere financial management, influencing the broader landscape of organizational effectiveness and competitiveness.

In the context of business estimation, understanding the business requirements entails recognizing the necessity

for accurate, timely, and comprehensive estimation of operating expenses. This involves engaging with business stakeholders to ascertain their specific needs and expectations regarding expense estimation. The ultimate objective is to align the analytical goals with the broader business objectives of achieving precise financial projections and optimizing resource allocation.

## PROBLEM SOLUTION FIT

**Table-1 : Components & Technologies:**

| S.No | Component | Description | Technology |
|---|---|---|---|
| 1. | User Interface | Web UI-IBM cognos | HTML |
| 2. | Application Logic-1 | Preparing the dataset | Python |
| 3. | Application Logic-2 | Data Exploration | IBM Cognos |
| 4. | Application Logic-3 | Data Visualization | IBM Cognos |
| 5. | Cloud Database | Database Service on Cloud | IBM DB2 |
| 6. | File Storage | File storage requirements | IBM Block Storage or Other Storage Service or Local File system |
| 7. | Infrastructure (Server / Cloud) | Application Deployment on Local System / Cloud<br>Local Server Configuration:<br>Cloud Server Configuration : | Local, Cloud Foundry, Kubernetes, etc. |

**Table-2: Application Characteristics:**

| S.No | Characteristics | Description | Technology |
|---|---|---|---|
| 1. | Open-Source Frameworks | For data pre-processing and data cleaning | Google Collaboratory, Jupyter notebook |
| 2. | Scalable Architecture | Supports various data sizes | IBM Cognos Analhytics |
| 3. | Availability | Supports all compactable web browser | Any web browser |
| 4. | Performance | Processing huge amount of data and make dashboards | IBM Cognos Analytics |

## REQUIREMENT ANALYSIS

### Functional requirements

Functional requirements define the specific functionalities and features that the system or project must deliver to meet the objectives and needs of its users. In the context of your data analytics project on operating expenses in the education sector, the following functional requirements may be considered:

User Authentication and Authorization:

Requirement: The system must have a secure user authentication mechanism.
Rationale: To ensure that only authorized users can access sensitive financial data.
Data Import and Integration:

Requirement: The system should allow for the import and integration of financial data from diverse sources.
Rationale: To ensure that the dataset is comprehensive and reflects the financial dynamics of various educational institutions.
Real-Time Data Updates:

Requirement: The system must support real-time updates for financial data.
Rationale: To provide users with the most current information for analysis and decision-making.
Customizable Filters:

Requirement: Users should be able to apply customizable filters for time, institution, and

expense categories.
Rationale: To enable users to tailor analyses based on specific criteria and focus on relevant aspects of operating expenses.
Interactive Data Visualization:

Requirement: The system should provide interactive data visualization tools.
Rationale: To enhance the user experience and allow stakeholders to explore and interpret financial data intuitively.
Query Performance Optimization:

Requirement: The system must optimize query performance for efficient data retrieval.
Rationale: To ensure that users can analyze data seamlessly without significant delays.
Cross-Filtering Capabilities:

Requirement: Implement cross-filtering to enable users to filter data in one dimension and see the impact on other dimensions.
Rationale: To facilitate a deeper exploration of relationships within the data.
Export and Reporting Functionality:

Requirement: Users should be able to export analysis results and generate reports.
Rationale: To allow for sharing insights with stakeholders and supporting documentation.
User Feedback Mechanism:

Requirement: The system should include a feedback mechanism for users to provide input on data analysis tools and features.
Rationale: To continuously improve the system based on user feedback.
Security Measures:

Requirement: Implement robust security measures to protect sensitive financial data.
Rationale: To safeguard against unauthorized access and ensure compliance with data privacy regulations.
Scalability:

Requirement: The system should be scalable to handle a growing volume of financial data.
Rationale: To accommodate increasing data volumes without sacrificing performance.
Documentation and Training:

Requirement: Provide comprehensive documentation on system usage and functionalities.
Rationale: To ensure that users can effectively navigate and utilize the data analytics platform.
Backup and Recovery Mechanism:

Requirement: Implement a robust backup and recovery mechanism for financial data.
Rationale: To ensure data integrity and availability in the event of system failures.
These functional requirements form the basis for the development and implementation of the data analytics project, ensuring that the system meets the needs of stakeholders and facilitates effective analysis of operating expenses in the education sector.

## Non-Functional requirements

Non-functional requirements define the characteristics or qualities that describe how a system should perform, rather than specific behaviors. They are crucial for ensuring that the system meets certain criteria in terms of performance, reliability, usability, and other aspects. Here are non-functional requirements for your data analytics project on operating expenses in the education sector:

Performance:

Requirement: The system should provide fast response times for data retrieval and analysis.
Rationale: To ensure that users can interact with the system efficiently and analyze data without significant delays.
Scalability:

Requirement: The system should be scalable to handle an increased number of users and a growing volume of data.
Rationale: To accommodate the potential growth of users and data without compromising performance.
Reliability:

Requirement: The system must be highly reliable, with minimal downtime or disruptions.
Rationale: To ensure that users can access the system consistently and rely on its availability for critical decision-making.
Availability:

Requirement: The system should be available 24/7, with planned downtime communicated in advance.
Rationale: To meet the needs of users in different time zones and ensure continuous access to the system.
Security:

Requirement: The system must adhere to industry-standard security protocols to protect sensitive financial data.
Rationale: To prevent unauthorized access, data breaches, and ensure compliance with data protection regulations.
Usability:

Requirement: The user interface should be intuitive and user-friendly, requiring minimal training for users to navigate.
Rationale: To enhance user adoption and ensure that stakeholders can easily interact with and interpret the data.
Compatibility:

Requirement: The system should be compatible with a variety of devices and browsers.
Rationale: To ensure accessibility for users on different platforms and devices.
Interoperability:

Requirement: The system should be able to integrate with other relevant tools or databases.
Rationale: To facilitate seamless data exchange and integration with existing systems within educational institutions.
Compliance:

Requirement: The system must comply with relevant data protection and privacy regulations.
Rationale: To avoid legal implications and ensure ethical handling of sensitive financial data.
Maintainability:

Requirement: The system should be designed for ease of maintenance and updates.
Rationale: To facilitate ongoing improvements, updates, and modifications to meet evolving user needs.
Documentation:

Requirement: Comprehensive documentation should be provided for system administrators and users.
Rationale: To aid in system understanding, troubleshooting, and support.
Data Backup and Recovery:

Requirement: Regular data backups should be performed, and a robust recovery mechanism should be in place.
Rationale: To safeguard against data loss and ensure the availability of historical financial data.
Audit Trail:
Requirement: The system should maintain an audit trail of user activities and changes to financial data.
Rationale: To support accountability, traceability, and compliance auditing.
By incorporating these non-functional requirements, you ensure that the data analytics project not only meets the functional needs of stakeholders but also delivers a reliable, secure.

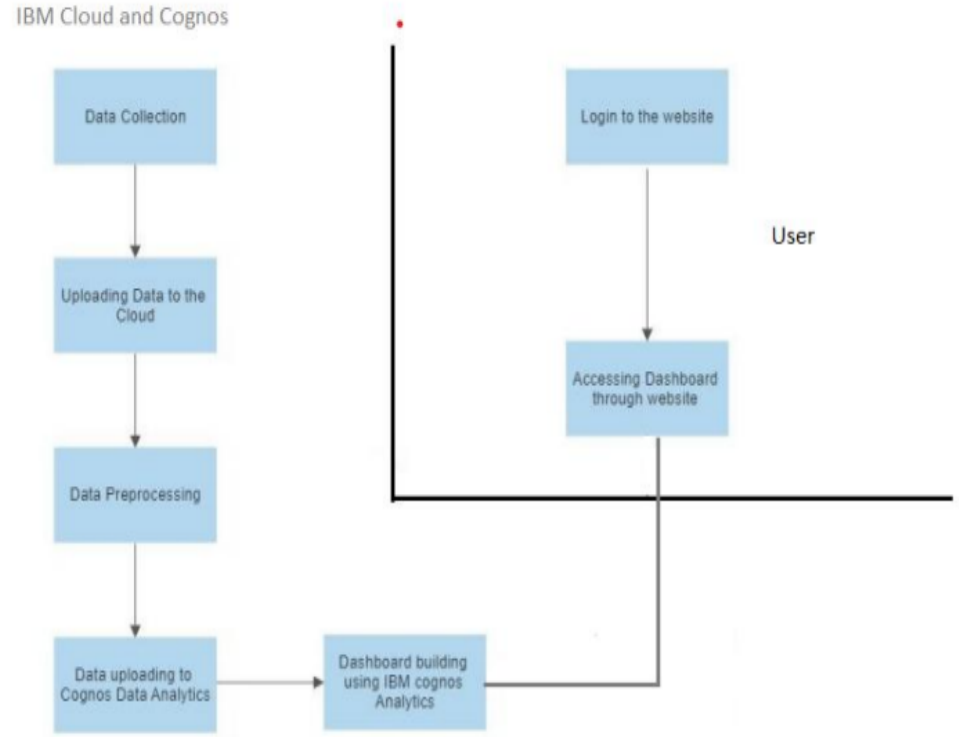
# PROJECT DESIGN

## Data Flow Diagrams

## FLOWS:

**Data Flow Diagrams:**

A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict the right amount of the system requirement graphically. It shows how data enters and leaves the system, what changes the information, and where data is stored.



## SOLUTION AND TECHNICAL ARCHITECTURE
## SOLUTION ARCHITECTURE

Certainly, let's dive deeper into each component of the solution architecture for estimating a business project using data analytics, including some additional considerations:

1. Data Sources:

   Identify and categorize data sources into internal and external. Internal sources may include historical project data, customer databases, and financial records, while external sources might encompass market trends, economic indicators, and industry reports.

2. Data Ingestion:

   Consider implementing real-time data ingestion for streaming data sources.

   Ensure data quality checks and error handling to handle incomplete or incorrect data.

3. Data Storage:

   Use data partitioning and indexing to optimize data retrieval performance.

   Implement data encryption and access controls to secure sensitive information.

4. Data Preprocessing:

   Apply data imputation techniques to handle missing values.

   Use outlier detection to identify and handle anomalies in the data.

5. Data Analysis and Modeling:

   Leverage exploratory data analysis (EDA) to gain insights into data patterns and relationships.

   Experiment with various machine learning algorithms and hyperparameter tuning to improve model accuracy.

6. Model Training and Evaluation:

- Employ cross-validation techniques for model evaluation to ensure robustness.
- Continuously monitor model performance and retrain models as needed.

7. Estimation Algorithms:

- Consider ensembling techniques like Random Forests or Gradient Boosting to combine multiple models for better accuracy.
- For time series forecasting, evaluate models like ARIMA, Prophet, or deep learning models.

8. Visualization and Reporting:

- Design interactive dashboards that allow users to explore and filter data.
- Provide scenario analysis to evaluate project outcomes under different assumptions.

9. Integration and Deployment:

- Implement APIs for easy integration with other systems and applications.
- Consider containerization (e.g., Docker) for portability and ease of deployment.

10. Monitoring and Maintenance:

- Set up automated alerts for anomalies in data or model performance.

- Implement version control for models and code to track changes over time.

11. Security and Compliance:

- Implement role-based access control to restrict data access based on user roles.

- Ensure compliance with GDPR, HIPAA, or industry-specific regulations, as applicable.

12. Scalability and Performance:

- Utilize serverless computing or auto-scaling capabilities to handle spikes in demand.

- Optimize queries and data pipelines for performance.

13. Documentation and Knowledge Sharing:

- Maintain a centralized knowledge base or wiki to document data sources, data dictionaries, model documentation, and best practices.

- Conduct regular training sessions to ensure that stakeholders understand how to use the system effectively.

14. Feedback Loop:

- Establish a process for collecting feedback from users and project managers and incorporate it into ongoing model improvement efforts.

- Track the accuracy of model predictions against actual project outcomes to refine estimation algorithms.

15. Disaster Recovery and Backup:

- Create a disaster recovery plan, including data backup and restoration procedures.

- Perform periodic disaster recovery drills to ensure readiness.

16. Budget and Cost Management:

- Implement cost monitoring tools and practices to optimize resource allocation and minimize expenses.

Remember that the specific tools, technologies, and architecture decisions will depend on your organization's preferences, expertise, and the nature of the projects you're estimating. Regularly review and update your solution architecture to stay current with evolving technologies and changing project requirements.

Data Collection

Data Analyst

Upload Data

Data Preprocessing

Analysis of Data through Visualization

Understanding the features influencing the Sales

Creating dashboard that can tell a meaningful sales strategy for future


Sales Team

## TECHNICAL ARCHITECTURE:

The technical architecture for estimating a business project using data analytics focuses on the underlying technology stack and infrastructure that supports the solution. Here's a detailed technical architecture for such a system:

1. Data Sources:

- Internal sources: Project history data, financial records, CRM systems, ERP systems.
- External sources: Market data, economic indicators, industry reports, government datasets, social media data.

2. Data Ingestion:

- Use data integration tools like Apache Nifi or Apache Kafka for real-time data ingestion.
- Implement batch processing for historical data using tools like Apache Spark or custom ETL (Extract, Transform, Load) scripts.

3. Data Storage:

- Use a data lake or data warehouse for storing structured and unstructured data.
- Consider cloud-based solutions like Amazon S3, Google Cloud Storage, or on-premises solutions like Hadoop HDFS.

4. Data Preprocessing:

- Implement data cleaning, transformation, and enrichment using technologies like Apache Spark or Pandas.
- Use data versioning to track changes and ensure reproducibility.

5. Data Analysis and Modeling:

- Utilize machine learning libraries and frameworks like Scikit-Learn, TensorFlow, or PyTorch for building and training predictive models.
- Experiment with Jupyter notebooks for model development and experimentation.

6. Model Training and Evaluation:

- Split data into training, validation, and test sets.
- Implement cross-validation and grid search to fine-tune model hyperparameters.
- Leverage automated machine learning (AutoML) platforms for model selection and tuning.

7. Estimation Algorithms:

- Depending on the nature of the project, use regression analysis, time series forecasting, or classification algorithms.
- Experiment with deep learning models for complex patterns in the data.

8. Visualization and Reporting:

- Create interactive dashboards and reports using visualization tools like Tableau, Power BI, or custom web-based interfaces.
- Embed charts and graphs into web applications for user-friendly access.

9. Integration and Deployment:

- Deploy the system on cloud infrastructure for scalability and reliability.
- Use containerization (e.g., Docker) for portability and microservices for modularity.
- Implement API endpoints for integration with other business applications.

10. Monitoring and Maintenance:

- Implement real-time monitoring and alerting for system health and performance.

- Use log aggregation and analysis tools (e.g., ELK Stack) to track system activities and diagnose issues.

11. Security and Compliance:

- Apply encryption for data at rest and in transit.

- Implement identity and access management (IAM) controls to restrict data access.

- Ensure compliance with data privacy regulations and audit trails for data access.

12. Scalability and Performance:

- Utilize auto-scaling in the cloud to handle varying workloads.

- Load balancing and caching can improve system performance.

13. Documentation and Knowledge Sharing:

- Maintain documentation for data sources, data transformations, and model details.

- Foster knowledge sharing through collaboration tools like Confluence or internal wikis.

14. Feedback Loop:

- Implement mechanisms to gather user feedback and track the accuracy of estimations compared to actual project outcomes.

- Continuously improve models based on real-world feedback.

15. Disaster Recovery and Backup:

- Regularly back up data and model artifacts.

- Develop a disaster recovery plan for system restoration in case of catastrophic failure.

16. Budget and Cost Management:

- Use cost monitoring tools to track and optimize cloud resource spending.

- Implement budget controls to manage expenses effectively.

This technical architecture serves as the foundation for building a robust and scalable system for estimating business projects using data analytics. It should be adjusted to match your organization's specific technology preferences, existing infrastructure, and project requirements. Regularly review and update the architecture to incorporate new technologies and best practices.

**Technical Architecture:**

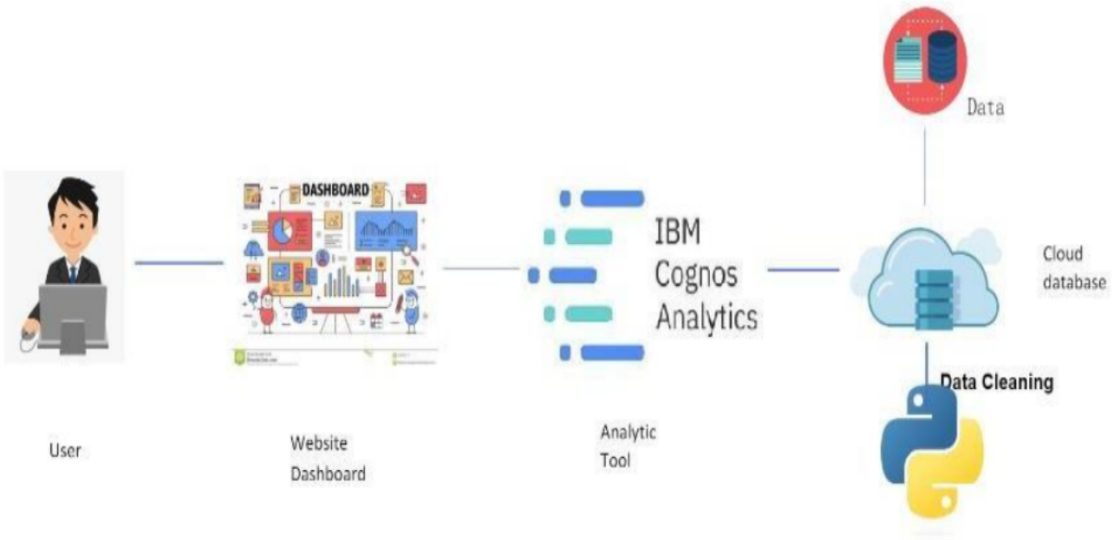


Table-1 :

Table-1 : Components & Technologies

Certainly, here's a table with components, descriptions, and corresponding technologies for the estimation of a business project using data analytics:

| S.No | Component | Description | Technology |
|---|---|---|---|
| 1 | Data Sources | Identify and collect data sources for project estimation. | Database Systems, APIs, External Datasets |
| 2 | Data Ingestion | Process of collecting and importing data into the system. | Apache Nifi, Apache Kafka, Custom ETL Scripts |
| 3 | Data Storage | Repository for storing both raw and processed data. | Data Lake, Data Warehouse (e.g., Amazon S3, Hadoop HDFS) |
| 4 | Data Preprocessing | Cleaning, transformation, and data quality improvement. | Apache Spark, Pandas, Data Versioning |
| 5 | Data Analysis and Modeling | Building predictive models for estimation. | Scikit-Learn, TensorFlow, PyTorch, Jupyter Notebooks |
| 6 | Model Training and Evaluation | Training and validating machine learning models. | Cross-Validation, Grid Search, AutoML Platforms |
| 7 | Estimation Algorithms | Algorithms for estimating project outcomes. | Regression, Time Series Forecasting, Deep Learning |
| 8 | Visualization and Reporting | Creating user-friendly interfaces for results. | Tableau, Power BI, Web-Based Dashboards, Chart Libraries |
| 9 | Integration and Deployment | Connecting the system with other applications. | Cloud Infrastructure, Docker, APIs |
| 10 | Monitoring and Maintenance | Continuous system health monitoring and upkeep. | Real-Time Monitoring, Log Aggregation, Alerting |

| | | | |
|---|---|---|---|
| 11 | Security and Compliance | Safeguarding data and ensuring regulatory compliance. | Data Encryption, Identity and Access Management (IAM), Compliance Tools |
| 12 | Scalability and Performance | Handling system growth and optimizing performance. | Cloud Auto-Scaling, Load Balancing, Caching |
| 13 | Documentation and Knowledge Sharing | Creating and sharing documentation. | Confluence, Internal Wikis, Collaboration Tools |
| 14 | Feedback Loop | Gathering user feedback and improving models. | Feedback Collection Mechanisms, Model Evaluation Metrics |
| 15 | Disaster Recovery and Backup | Planning for system recovery in case of failure. | Backup Strategies, Disaster Recovery Plan, Restore Procedures |
| 16 | Budget and Cost Management | Tracking and managing project costs. | Cost Monitoring Tools, Budget Controls, Cloud Resource Optimization |

This table provides an organized overview of the key components and the technologies associated with each component in the architecture for business project estimation using data analytics.

**Table-2: Application Characteristics:**

Certainly, here's a table with application characteristics, descriptions, and the corresponding technologies for the estimation of a business project using data analytics:

| S.No | Characteristics | Description | Technology |
|---|---|---|---|
| 1 | Scalability | Ability to handle growing data and user loads. | Cloud Services (e.g., AWS, Azure), Auto-Scaling |
| 2 | Real-time Processing | Handling data streams and providing real-time insights. | Apache Kafka, Stream Processing Frameworks, Pub-Sub Systems |
| 3 | Data Security | Ensuring data protection and access control. | Encryption, IAM (Identity and Access Management), Auditing |

| | | | |
|---|---|---|---|
| 4 | Performance Optimization | Enhancing system speed and resource utilization. | Load Balancing, Caching, Query Optimization |
| 5 | User-Friendly Interface | Providing intuitive dashboards and reports for users. | Web Development Frameworks, UI/UX Design Tools |
| 6 | Collaboration Support | Enabling teams to work together and share insights. | Collaboration Tools, Communication Platforms |
| 7 | Data Privacy Compliance | Meeting regulatory requirements for data protection. | Compliance Tools, Data Masking, Policy Enforcement |
| 8 | Machine Learning Capabilities | Integrating AI/ML models for advanced estimation. | ML Frameworks (e.g., TensorFlow, Scikit-Learn), Model Deployment Platforms |
| 9 | Data Backup and Recovery | Ensuring data integrity and recovery in case of failure. | Backup Solutions, Disaster Recovery Planning |
| 10 | Cost Management | Monitoring and managing project-related expenses. | Cost Monitoring Tools, Budgeting Strategies |
| 11 | Mobile Access | Supporting mobile devices for remote access. | Mobile App Development, Responsive Web Design |
| 12 | Multilingual Support | Handling different languages for global use. | Localization Libraries, Multilingual Data Management |
| 13 | Analytics Insights | Providing data-driven insights for informed decision-making. | Advanced Analytics Tools, Statistical Analysis Libraries |
| 14 | Data Versioning | Tracking changes to data for reproducibility. | Version Control Systems, Data Versioning Techniques |
| 15 | System Monitoring | Continuous health and performance monitoring. | Monitoring Tools, Log Analysis, Anomaly Detection |

This table summarizes the various application characteristics and the associated technologies that are important for a system designed to estimate business projects using data analytics. These characteristics contribute to the system's effectiveness and usability in a real-world business environment.

## USER STORIES

User stories can help define the specific requirements and tasks for different user types in the context of estimating a business project using data analytics. Here's a table with user stories, user types, functional requirements, user story numbers, user tasks/stories, acceptance criteria, and priorities:

| User Type | Functional Requirement | User Story Number | User Task/Story | Acceptance Criteria | Team member | Priority |
|---|---|---|---|---|---|---|
| Business Stakeholder | Scenario Analysis | US-013 | As a business stakeholder, I want to perform scenario analysis to evaluate project outcomes under different assumptions. | 1. Ability to set and vary input parameters. 2. The system provides results for different scenarios. | Harinieswari V | High |
| Data Analyst | Data Version Control | US-014 | As a data analyst, I need to track changes in data for reproducibility and audit purposes. | 1. Data versioning is implemented. 2. Historical data changes are traceable. | Srimathi T | High |
| Data Scientist | Advanced Analytics | US-015 | As a data scientist, I want access to advanced analytics tools for in-depth data exploration. | 1. Integration with advanced analytics libraries (e.g., R) or specialized tools (e.g., data mining software). 2. Ability to perform advanced statistical analysis. | Vaishnavi R | Medium |

| | | | | | | |
|---|---|---|---|---|---|---|
| System Administrator | Load Balancing | US-016 | As a system administrator, I need to implement load balancing to distribute user requests efficiently. | 1. Load balancing is configured to evenly distribute traffic. 2. System performance is improved during peak loads. | Harinieswari V | Medium |
| Data Privacy Officer | Data Masking | US-017 | As a data privacy officer, I need to implement data masking to protect sensitive information. | 1. Sensitive data is masked or anonymized in reports and dashboards. 2. Privacy policies are enforced. | Srimathi T | Medium |
| Project Manager | Model Reevaluation | US-018 | As a project manager, I need to periodically reevaluate models with new data for accuracy. | 1. Models are retrained with new data at regular intervals. 2. Model accuracy is improved over time. | Vaishnavi R | Low |
| Data Scientist | Anomaly Detection | US-019 | As a data scientist, I need to implement anomaly detection to identify unusual patterns in project data. | 1. Anomaly detection algorithms are integrated into the system. 2. Alerts are generated when anomalies are detected. | Harinieswari V | Low |

| | | | | | | |
|---|---|---|---|---|---|---|
| Business Stakeholder | Collaborative Reporting | US-020 | As a business stakeholder, I want to collaborate on reports and insights with team members. | 1. Collaborative reporting and annotation features are available. 2. Multiple users can contribute to reports. | Srimathi T | Low |
| System Administrator | System Health Monitoring | US-021 | As a system administrator, I need to continuously monitor system health for proactive issue resolution. | 1. Real-time system health monitoring tools are implemented. 2. Automated alerts are generated for critical issues. | Vaishnavi R | Low |
| Data Analyst | Statistical Analysis | US-022 | As a data analyst, I want access to statistical analysis libraries to conduct hypothesis testing and inferential statistics. | 1. Integration with statistical analysis libraries (e.g., StatsModels, SAS). 2. Ability to perform hypothesis testing and generate statistical reports. | Harinieswari V | Low |

These additional user stories cover a broader range of functionalities and address the needs of different user types involved in the estimation of business projects using data analytics. The priorities can be adjusted based on your project's specific requirements and constraints.

| Sprint | Functional Requirement | User Story Number | User Task/Story | Acceptance Criteria | Priority |
|---|---|---|---|---|---|
| Sprint 1 | Data Import | US-001 | As a project manager, I want to upload historical project data for analysis. | 1. I can upload data in common formats (CSV, Excel). 2. Data is correctly processed and stored. | High |
| Sprint 1 | Data Preprocessing | US-002 | As a data analyst, I need to clean and preprocess raw project data. | 1. Data cleaning and transformation are automated. 2. Data quality improves after preprocessing. | High |
| Sprint 1 | Model Development | US-003 | As a data scientist, I want to create predictive models for project estimation. | 1. I can experiment with various machine learning algorithms. 2. Models can be trained and tested. | High |
| Sprint 1 | Results Visualization | US-004 | As a business stakeholder, I need to view estimation results in a dashboard. | 1. Access to a user-friendly dashboard. 2. Charts and visualizations display project estimates. | High |

| | | | | | |
|---|---|---|---|---|---|
| Sprint 1 | Integration & Deployment | US-005 | As a system administrator, I need to deploy the system for end-users. | 1. The system is deployed on the specified platform (e.g., cloud or on-premises). 2. Integration with other business systems is successful. | High |
| Sprint 2 | Security & Compliance | US-006 | As a data privacy officer, I must ensure data security and regulatory compliance. | 1. Data is encrypted and access-controlled. 2. Compliance with relevant data protection regulations. | High |
| Sprint 2 | Model Feedback | US-007 | As a project manager, I want to provide feedback on model performance. | 1. A feedback mechanism is available. 2. Feedback is used to improve estimation models. | Medium |
| Sprint 2 | Real-time Processing | US-008 | As a data analyst, I need to process real-time project data for immediate insights. | 1. The system handles data streams in real-time. 2. Real-time insights are generated. | Medium |
| Sprint 2 | Mobile Access | US-009 | As a business stakeholder, I want to access project estimation results on my mobile device. | 1. The system supports mobile access via responsive design or a mobile app. | Medium |

| | | | | | |
|---|---|---|---|---|---|
| Sprint 2 | Disaster Recovery | US-010 | As a system administrator, I must plan for disaster recovery. | 1. Disaster recovery procedures are documented. 2. Regular backup and recovery tests are conducted. | Medium |
| Sprint 3 | Cost Monitoring | US-011 | As a project manager, I need to monitor and control the project-related expenses. | 1. Cost monitoring tools are integrated into the system. 2. Budget controls are in place. | Low |
| Sprint 3 | Multilingual Support | US-012 | As a data scientist, I need to work with data in multiple languages for global projects. | 1. The system supports data in various languages. 2. Localization libraries are used for multilingual data management. | Low |

These user stories are divided into sprints to represent the iterative nature of Agile development. The priorities can be adjusted based on project needs and resource constraints within each sprint

**CODING AND SOLUTION:**

**HTML CODE:**

```
<!DOCTYPE html>
<html lang="en" >
  <head>
    <meta charset="UTF-8" />
    <meta http-equiv="X-UA-Compatible" content="IE=edge" />
    <meta name="viewport" content="width=device-width, initial-scale=1.0" />
    <title>Global Sales Data Analysis - IBM</title>
    <script src="https://cdn.tailwindcss.com"></script>
  </head>
<body class="h-screen overflow-hidden scroll-smooth bg-gray-100" background="https://png.pngtree.com/thumb_back/fh260/background/20211118/pngtree-technology-round-dashboard-image_908915.jpg">
   <header
        class="fixed top-0 p-4 bg-white border-b w-full shadow-md flex gap-8 justify-between items-center">
     <h1 class="font-bold text-lg">Global Sales Data Analytics</h1>
    </header>
    <center>
    <section id="report" class="h-screen p-5 pt-20">
       <iframe

src="https://us1.ca.analytics.ibm.com/bi/?perspective=dashboard&amp;pathRef=.my_folders%2FGlobalSalesAnalytics&amp;closeWindowOnLastView=true&amp;ui_appbar=false&amp;ui_navbar=false&amp;shareMode=embedded&amp;action=view&amp;mode=dashboard&amp;subView=model000001848baeeca4_00000000"
       width="1000px"
       height="600px"
       frameborder="0"
       gesture="media"
```

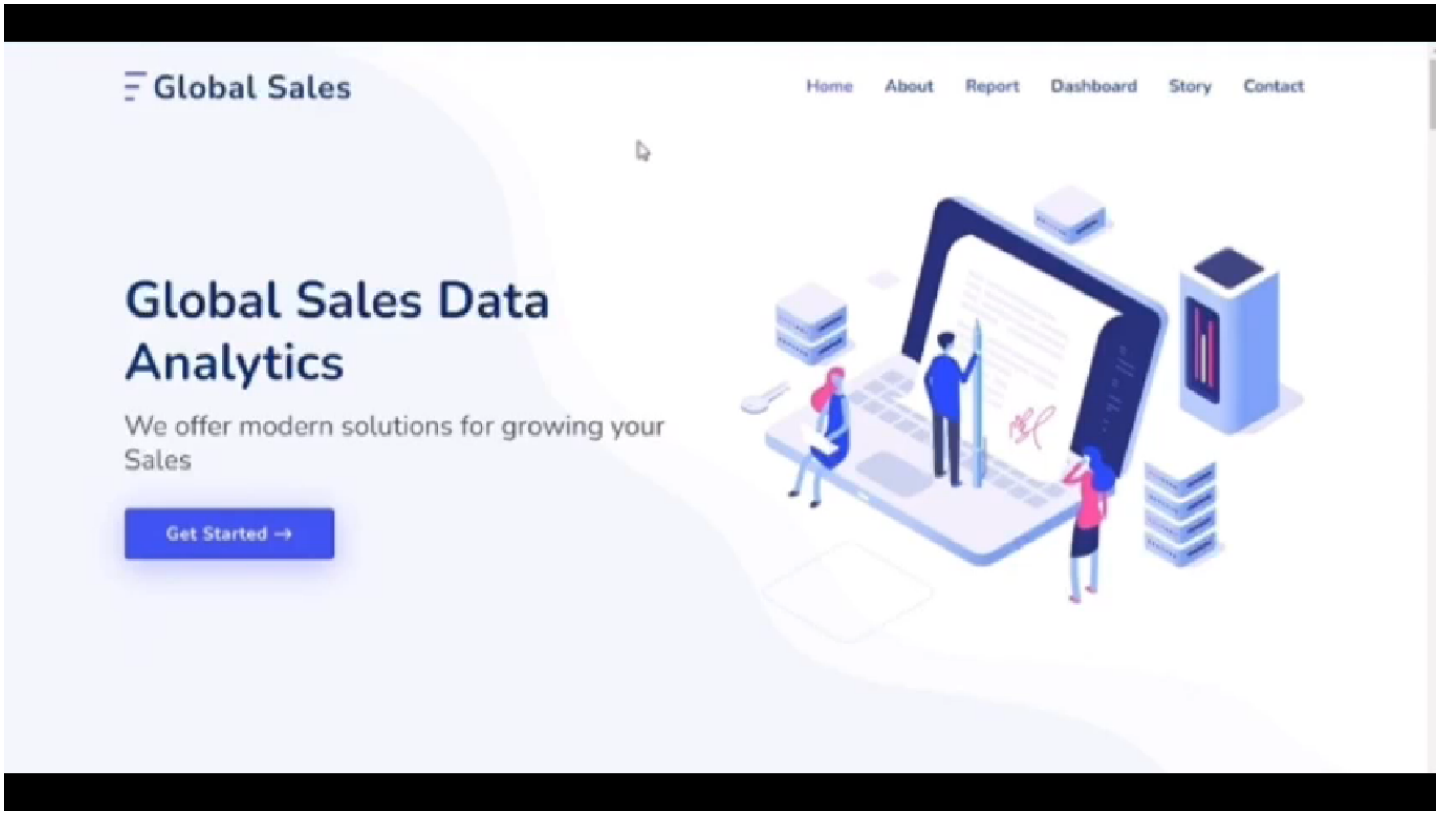
```
        allow="encrypted-media"
        allowfullscreen=""
        class="border"
      ></iframe>
    </section>
    </center>
  </body>
  <script>
    const links = document.querySelectorAll("a.link");
    links.forEach((el) =>
      el.addEventListener("click", (e) => {
        e.preventDefault();
        document
          .getElementById(el.getAttribute("data-href"))
          .scrollIntoView({ behavior: "smooth" });
        const currActive = document.querySelector("a.link.active");
        currActive?.classList.remove("active");
        currActive?.classList.remove("bg-blue-600");
        currActive?.classList.remove("text-white");
        el.classList.add("active");
        el.classList.add("bg-blue-600");
        el.classList.add("text-white");
      })
    );
  </script>
</html>
```

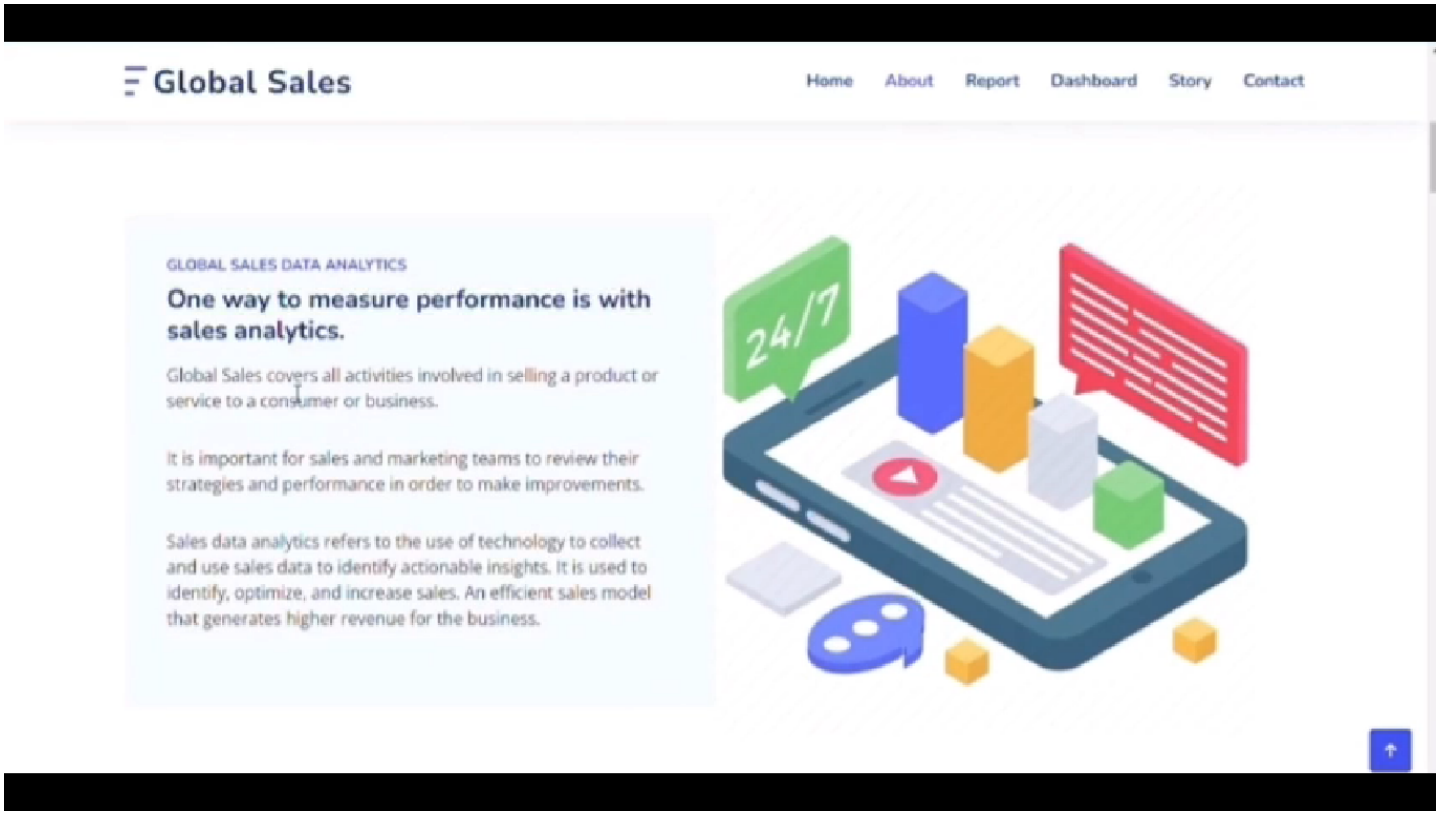
**FLASK CODE:**

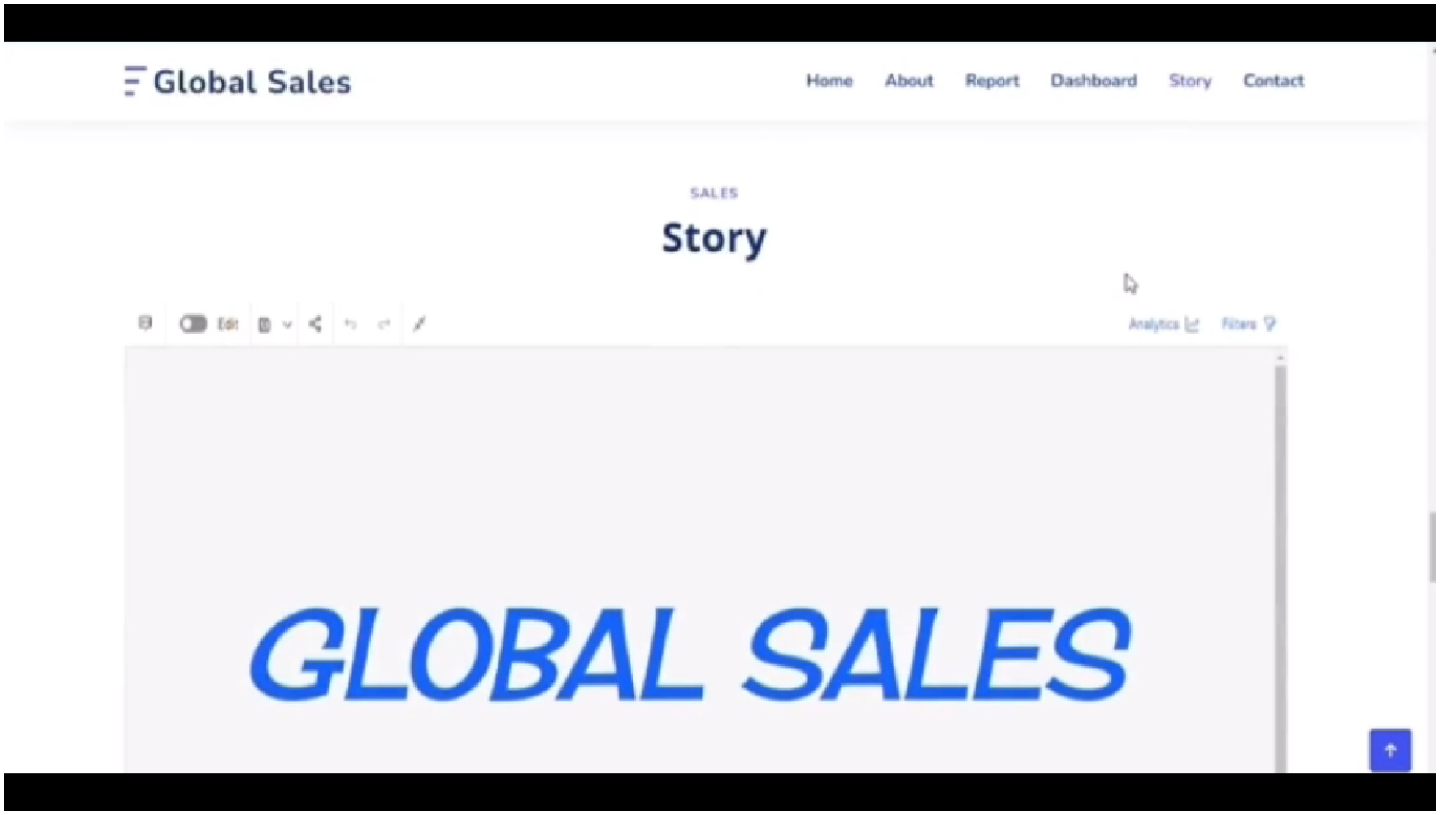
```
from flask import Flask, render_template

app = Flask(__name__)
```

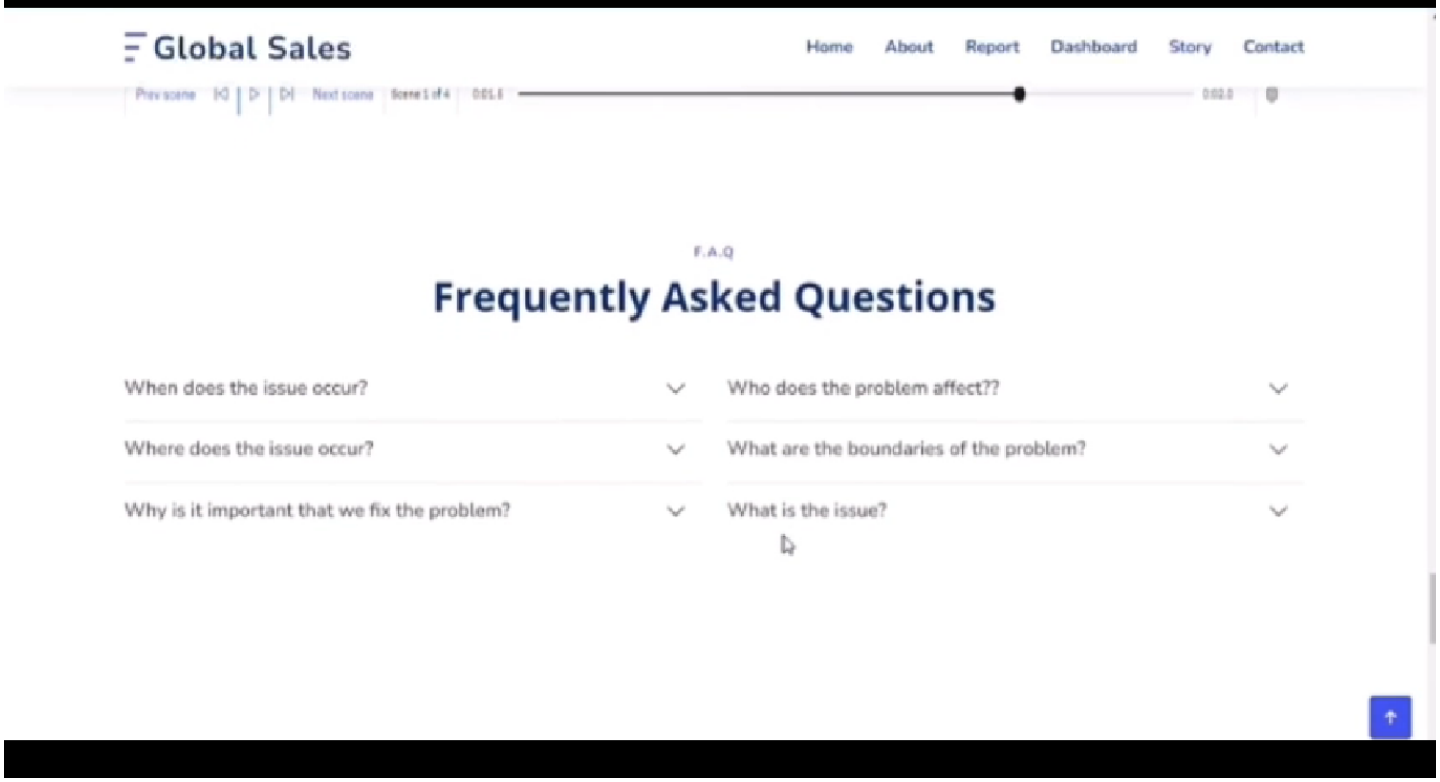
```
@app.route("/")
def index():
    return render_template("index.html")

if __name__ =="__main__":
    app.run
```

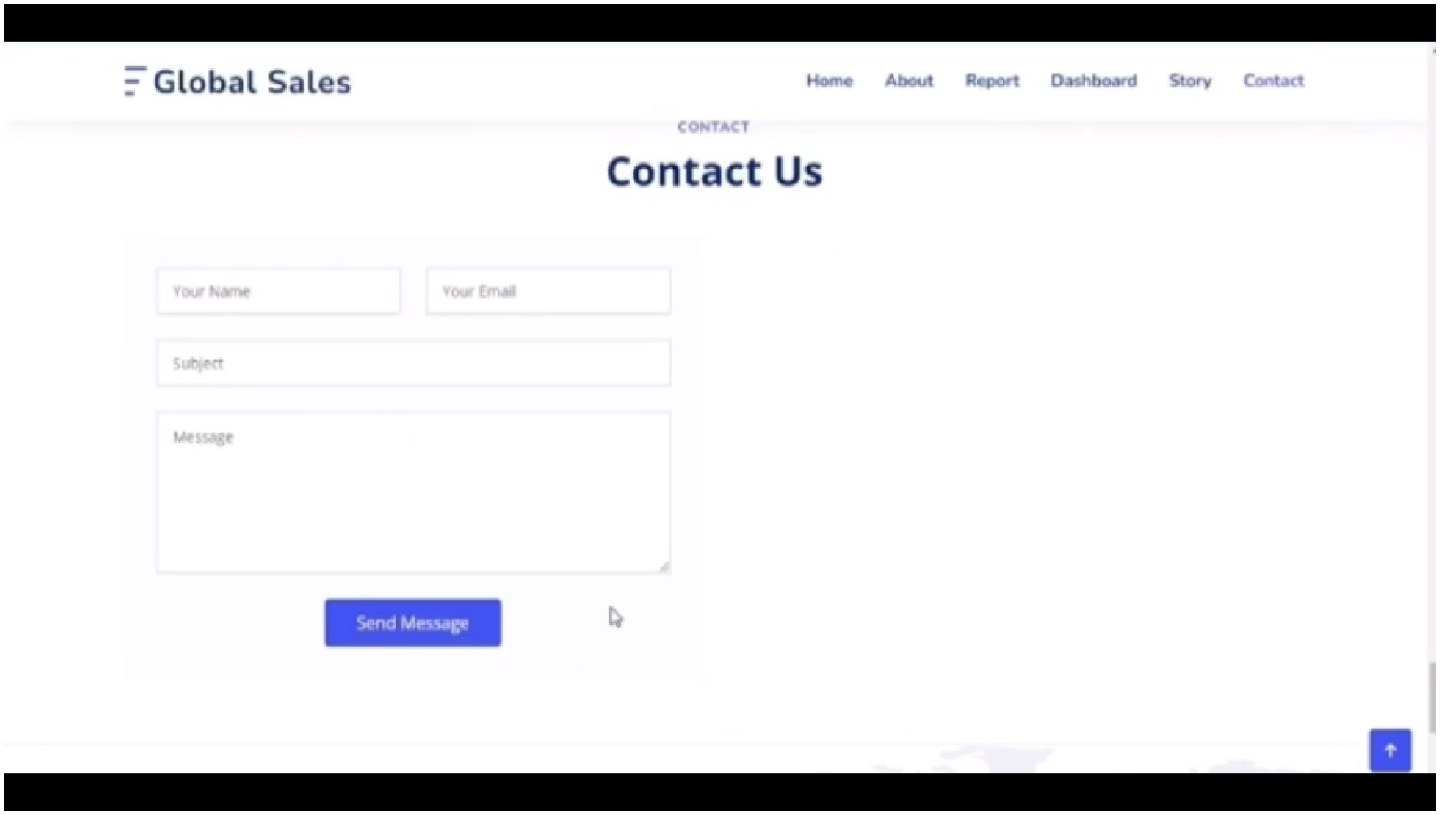
## OUTPUT SCREENSHOTS:

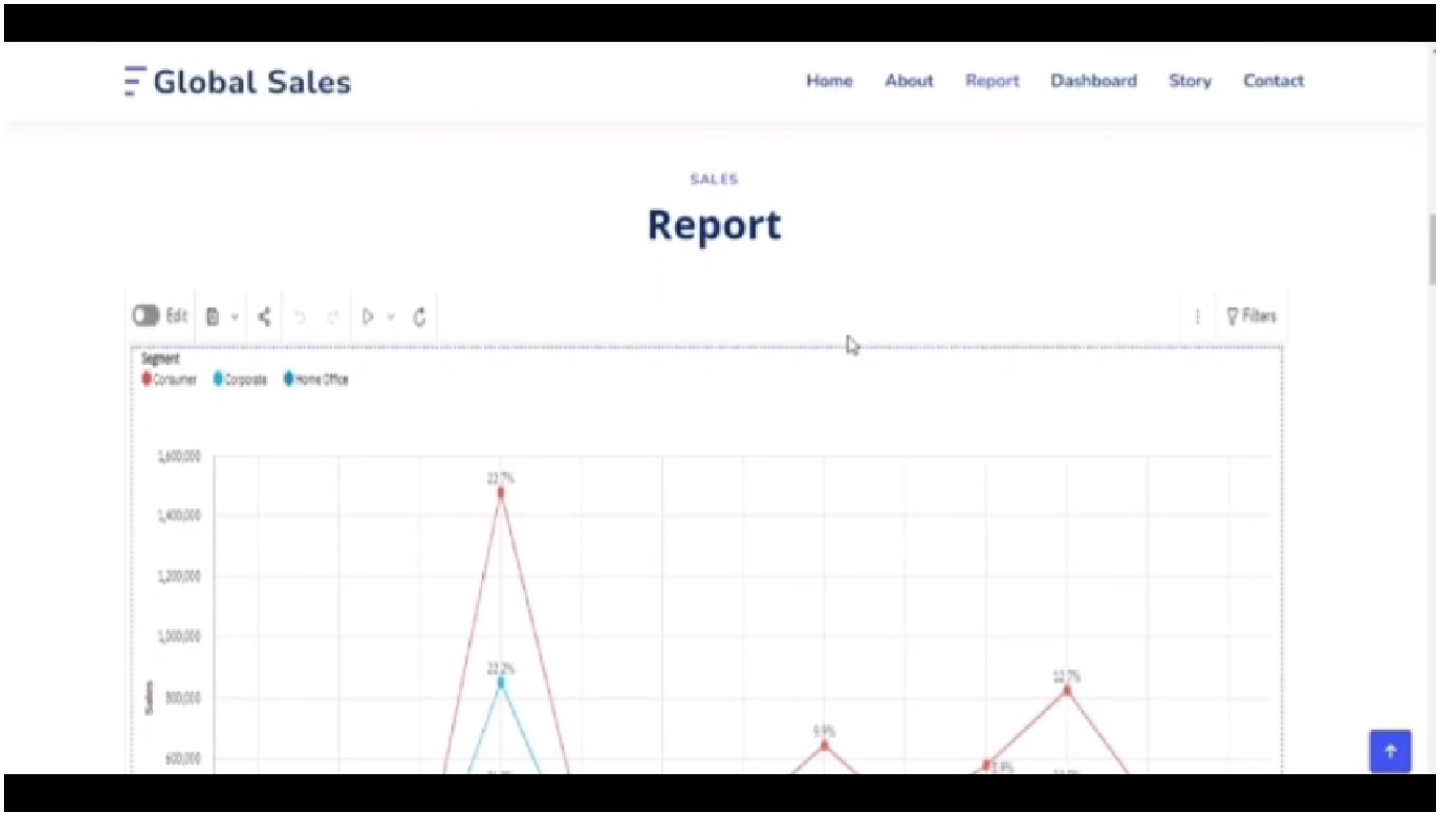



Global Sales
Home
About
Report
Dashboard
Story
Contact
SALES
Story
Edit
Analytics
Filters
GLOBAL SALES

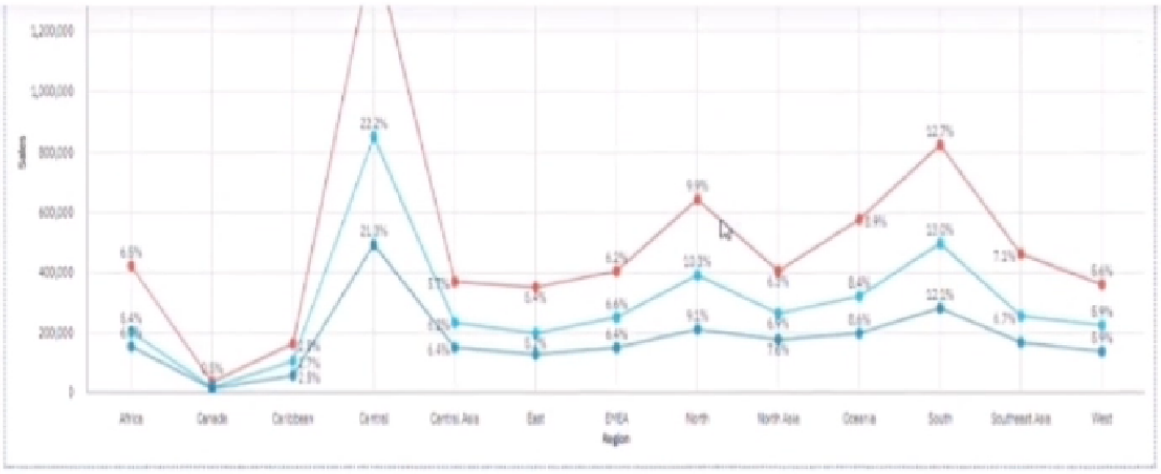

Global Sales
Home
About
Report
Dashboard
Story
Contact
F.A.Q
Frequently Asked Questions
When does the issue occur?
Who does the problem affect??
Where does the issue occur?
What are the boundaries of the problem?
Why is it important that we fix the problem?
What is the issue?


Global Sales
Home
About
Report
Dashboard
Story
Contact
CONTACT
Contact Us
Your Name
Your Email
Subject
Message
Send Message


Global Sales
Home
About
Report
Dashboard
Story
Contact
SALES
Report
Edit
Filters
Segment
Consumer
Corporate
Home Office
1,600,000
1,400,000
1,200,000
1,000,000
800,000
600,000
Sales
22.7%
22.2%
5.9%
11.9%
12.7%

Global Sales

Home About Report Dashboard Story Contact




Global Sales

Home About Report Dashboard Story Contact

## Report

sorting and organization of data, while analytics derive insights from that data and often influence business decisions.

## Stroy

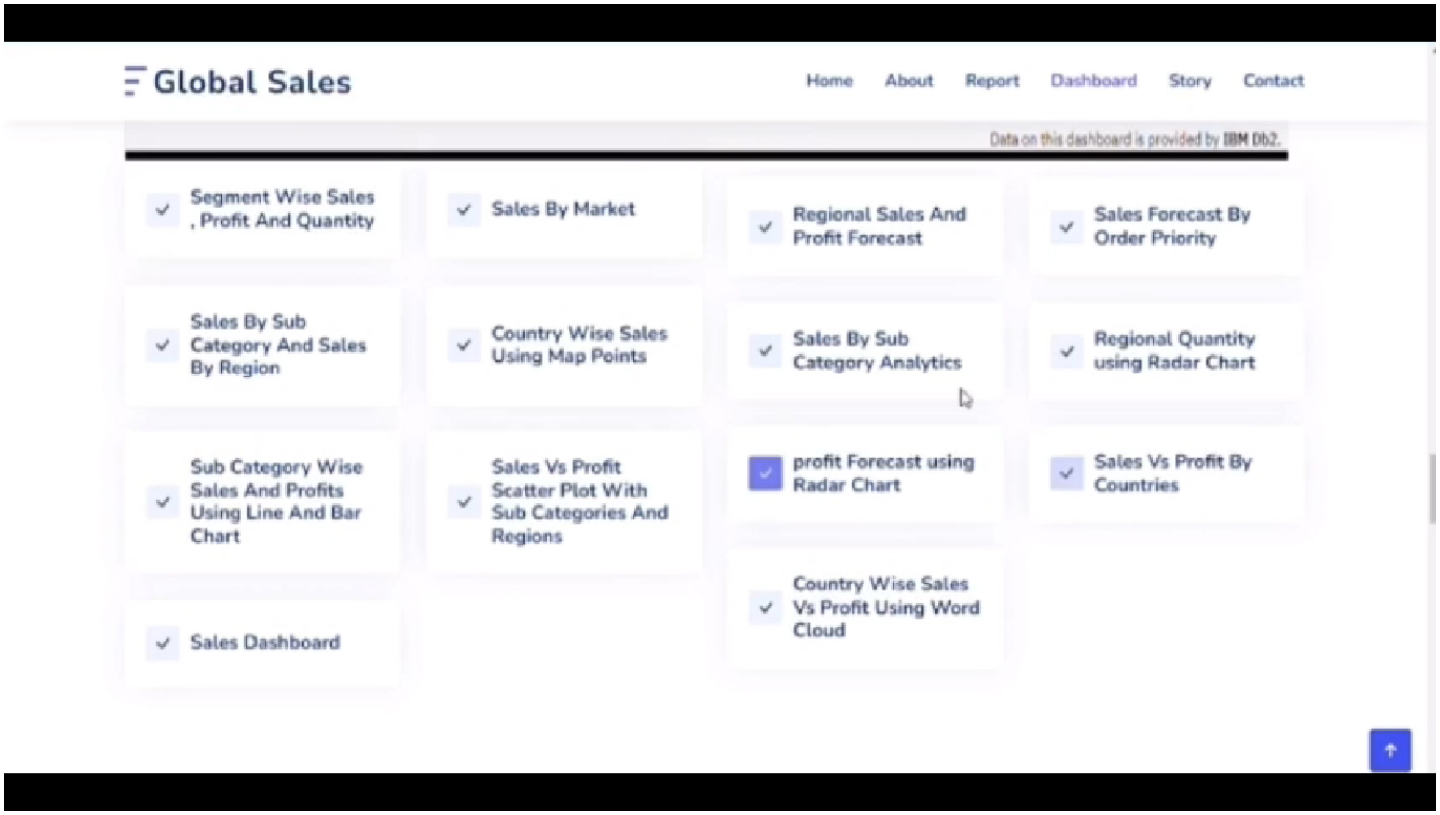
a methodology for communicating information, tailored to a specific audience, with a compelling narrative.

## Dashboard

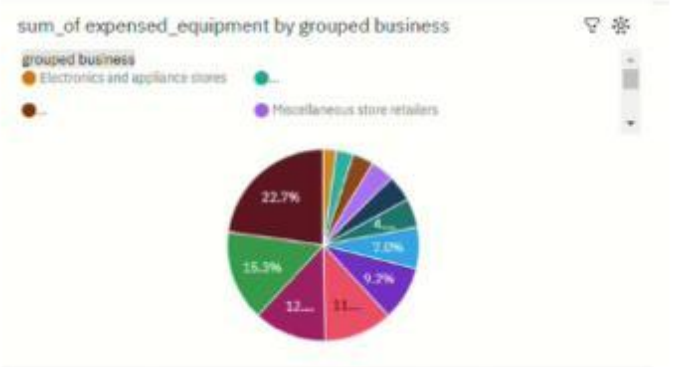
a tool used to multi-task, organize, visualize, analyze, and track data.View automatically updated data with interactive charts, graphs and tables.

SALES

# Dashboard

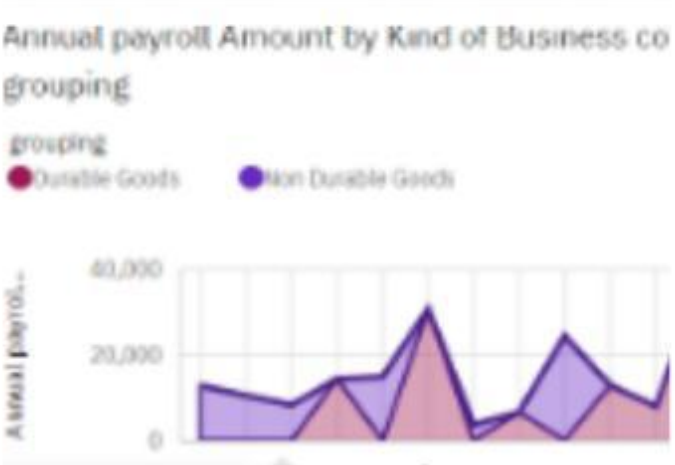

Global Sales
Home
About
Report
Dashboard
Story
Contact
Data on this dashboard is provided by IBM Db2.
Segment Wise Sales , Profit And Quantity
Sales By Market
Regional Sales And Profit Forecast
Sales Forecast By Order Priority
Sales By Sub Category And Sales By Region
Country Wise Sales Using Map Points
Sales By Sub Category Analytics
Regional Quantity using Radar Chart
Sub Category Wise Sales And Profits Using Line And Bar Chart
Sales Vs Profit Scatter Plot With Sub Categories And Regions
profit Forecast using Radar Chart
Sales Vs Profit By Countries
Country Wise Sales Vs Profit Using Word Cloud
Sales Dashboard

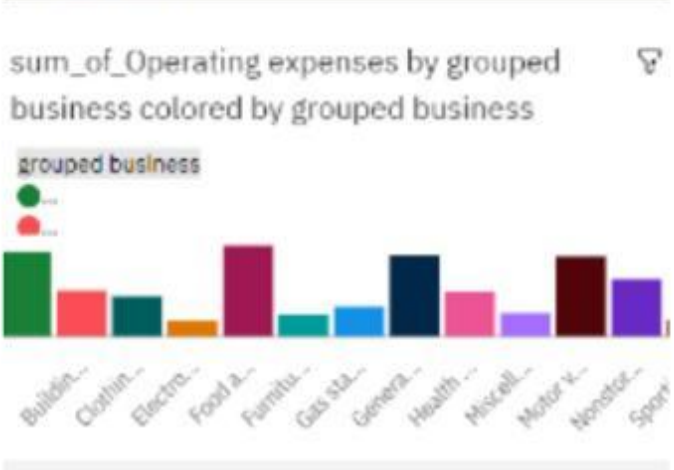

12_sum_of_expenses for Kind of Business hierarchy
Kind of Business
Apparel, Piece Goods, and Notions
Electrical Goods
Miscellaneous R...
Lumber & O...
Metals and ...
Paper and P...


sum_of expensed_equipment by grouped business
grouped business
Electronics and appliance stores
Miscellaneous store retailers
22.7%
15.3%
12....
11....
9.2%
7.0%


sum_of expensed_equipment by grouped business
grouped business
Electronics and appliance stores


sum_of_Operating expenses by grouped business colored by grouped business
sum_of_Expensed_purchases_of_software by grouped business
grouped business
Miscellaneous store retailers
Gas stations
Electronics and appliance stores
General merchandise stores
560 | 2...
196 |...
467 | ...
363...
2...


Annual payroll Amount by Kind of Business co
grouping
grouping
Durable Goods
Non Durable Goods
40,000
20,000
0


sum_of_annual_payroll_amount by grouped business colored by grouped business
sum_of_Operating expenses by grouped business colored by grouped business
grouped business
Buildin...
Clothin...
Electro...
Food a...
Furnitu...
Gas sta...
Genera...
Health ...
Miscell...
Motor v...
Nonstor...
Sport

# TESTING

CrimeVision CrimeClassification

Pass 80 Fail 0 Blocked 0 Query 0 80 / 80 100%

## Testing

| | | number<br>tester<br>date<br>build | 1.1<br>anyone<br>20 May 2023 | 2<br>anyone<br>20 May 2023 |
|---|---|---|---|---|
| 0001 | !pip install pandas<br><br>import pandas as pd<br>import numpy as np<br>import matplotlib.pyplot as plt<br>import seaborn as sns<br>import plotly.express as px<br>import os | | ✔ | ✔ |
| 0002 | <br>import tensorflow as tf<br>from tensorflow.keras.preprocessing import image_dataset_from_directory<br><br>from tensorflow.keras.applications import DenseNet121<br>f | | ✔ | ✔ |
| 0003 | from sklearn.preprocessing import LabelBinarizer<br>from tensorflow.keras.layers import Dense, GlobalAveragePooling2D, Dropout, MaxPooling2D, Conv2D<br>from tensorflow.keras.models import Sequential | | ✔ | ✔ |
| 0004 | <br>from IPython.display import clear_output<br>import warnings | | ✔ | ✔ |
| 0005 | warnings.filterwarnings("ignore") | | ✔ | ✔ |
| 0006 | train_dir="/content/NewTest"<br>test_dir="/content/NewTest" | | ✔ | ✔ |
| 0007 | <br>SEED=12<br>IMG_HEIGHT=64<br>IMG_WIDTH=64<br>BATCH_SIZE=128<br>EPOCHS=5<br>LR=0.00003 | | ✔ | ✔ |
| 0008 | crime_types=os.listdir(train_dir) | | ✔ | ✔ |

| | | | | |
|---|---|---|---|---|
| | LR=0.00003 | | | |
| 0008 | crime_types=os.listdir(train_dir)<br>n=len(crime_types)<br>print("Numeber of crime categories:",n) | | ✔ | ✔ |
| 0009 | crimes={}<br>train=test=0 | | ✔ | ✔ |
| 0010 | for clss in crime_types:<br>num=len(os.listdir(os.path.join(train_dir,clss))) | | ✔ | ✔ |
| 0011 | train+=num<br>test+=len(os.listdir(os.path.join(train_dir,clss)))<br><br>crimes[clss]=num | | ✔ | ✔ |
| 0012 | plt.figure(figsize=(8, 5))<br>plt.pie(x=np.array([train,test]),autopct="%.1f%%", explode=[0.1, 0.1], labels=["Training Data", "Test Data"], pctdistance=0.5, colors=['yellow','green']) | | ✔ | ✔ |
| 0013 | plt.title("Train and Test Images ", fontsize=18); | | ✔ | ✔ |
| 0014 | plt.figure(figsize=(15,5))<br><br>plt.bar(list(crimes.keys()), list(crimes.values()), width-0.4, align="center", edgecolor-['red'], color-["orange"]) | | ✔ | ✔ |
| 0015 | plt.xticks(rotation-90)<br><br>plt.xlabel("Reported Crimes") plt.ylabel("amber of Reported Crimes")<br><br>plt.show() | | ✔ | ✔ |
| 0016 | train_set=image_dataset_from_directory(<br><br>train_dir,<br><br>label_mode="categorical", | | ✔ | ✔ |
| 0017 | batch_size=BATCH_SIZE,<br><br>image_size=IMG_SHAPE,<br><br>shuffle=True, | | ✔ | ✔ |
| 0018 | batch_size=BATCH_SIZE,<br><br>image_size=IMG_SHAPE,<br><br>shuffle=True, | | ✔ | ✔ |
| 0019 | val_set=image_dataset_from_directory(<br><br>train_dir,<br><br>label_mode="categorical", | | ✔ | ✔ |
| 0020 | <br>batch_size=BATCH_SIZE,<br><br>image_size-IMG_SHAPE, | | ✔ | ✔ |

| | | | |
|---|---|---|---|
| | test_dir,<br>label_mode="categorical", class_names=None, | | |
| 0023 | batch_size=BATCH_SIZE,<br>image_size=IMG_SHAPE,<br>shuffle=False,<br>seed-seed,) | ✔ | ✔ |
| 0024 | def transfer learning():<br>base_model-DenseNet121(include_top=False, input_shape=INPUT_SHAPE, weights="imagenet")<br>thr=149 | ✔ | ✔ |
| 0025 | for layers in base_model.layers[:thr]: layers.trainable=False<br>for layers in base_model.layers[thr:]: layers.trainable=False<br>return base_model | ✔ | ✔ |
| 0026 | def create_model(): model-Sequential()<br>base_model transfer_learning() model.add(base_model)<br>model.add(GlobalAveragePooling2D())<br>model.add(Dense (256, activation="relu")) model.add(Dropout (8.2)) | ✔ | ✔ |
| 0027 | model.add(Dense (n, activation="softmax"))<br>model.summary()<br>return model | ✔ | ✔ |
| 0028 | model=create_model()<br>model.compile(optimizer="adam",<br>loss='categorical_crossentropy', metrics = ['accuracy']) | ✔ | ✔ |
| 0029 | from tensorflow.keras.models import load_model<br>model.load_weights('crime.h5') | ✔ | ✔ |
| 0030 | y_pred=model.predict(test_set) | ✔ | ✔ |
| 0031 | img<br>=<br>image.load_img('/content/Test/Robbery/Robbery048_x264_0.png',target_size=(64,64)) # Reading image<br>x = image.img_to_array(img) # Converting image into array | ✔ | ✔ |
| 0032 | x = np.expand_dims(x,axis=0) # expanding Dimensions<br>pred = np.argmax(model.predict(x)) # Predicting the higher probablity index | ✔ | ✔ |
| 0033 | img<br>image.load_img('/content/Test/Burglary/Burglary005_x264_0.png',target_size=(64,64)) # Reading image<br>X = image.img_to_array(img) # Converting image into array<br>x = np.expand_dims (x,axis=0) # expanding Dimensions | ✔ | ✔ |


| | | | |
|---|---|---|---|
| | test_dir,<br>label_mode="categorical", class_names=None, | | |
| 0023 | batch_size=BATCH_SIZE,<br>image_size=IMG_SHAPE,<br>shuffle=False,<br>seed-seed,) | ✔ | ✔ |
| 0024 | def transfer learning():<br>base_model-DenseNet121(include_top=False, input_shape=INPUT_SHAPE, weights="imagenet")<br>thr=149 | ✔ | ✔ |
| 0025 | for layers in base_model.layers[:thr]: layers.trainable=False<br>for layers in base_model.layers[thr:]: layers.trainable=False<br>return base_model | ✔ | ✔ |
| 0026 | def create_model(): model-Sequential()<br>base_model transfer_learning() model.add(base_model)<br>model.add(GlobalAveragePooling2D())<br>model.add(Dense (256, activation="relu")) model.add(Dropout (8.2)) | ✔ | ✔ |
| 0027 | model.add(Dense (n, activation="softmax"))<br>model.summary()<br>return model | ✔ | ✔ |
| 0028 | model=create_model()<br>model.compile(optimizer="adam",<br>loss='categorical_crossentropy', metrics = ['accuracy']) | ✔ | ✔ |
| 0029 | from tensorflow.keras.models import load_model<br>model.load_weights('crime.h5') | ✔ | ✔ |
| 0030 | y_pred=model.predict(test_set) | ✔ | ✔ |
| 0031 | img<br>=<br>image.load_img('/content/Test/Robbery/Robbery048_x264_0.png',target_size=(64,64)) # Reading image<br>x = image.img_to_array(img) # Converting image into array | ✔ | ✔ |
| 0032 | x = np.expand_dims(x,axis=0) # expanding Dimensions<br>pred = np.argmax(model.predict(x)) # Predicting the higher probablity index | ✔ | ✔ |
| 0033 | img<br>image.load_img('/content/Test/Burglary/Burglary005_x264_0.png',target_size=(64,64)) # Reading image<br>X = image.img_to_array(img) # Converting image into array | ✔ | ✔ |

| | | | |
|---|---|---|---|
| | from tensorflow.keras.models import load_model<br>model.load_weights('crime.h5') | | |
| 0030 | y_pred=model.predict(test_set) | ✔ | ✔ |
| 0031 | img<br>=<br>image.load_img('/content/Test/Robbery/Robbery048_x264_0.png',target_size=(64,64)) # Reading image<br>x = image.img_to_array(img) # Converting image into array | ✔ | ✔ |
| 0032 | x = np.expand_dims(x,axis=0) # expanding Dimensions<br>pred = np.argmax(model.predict(x)) # Predicting the higher probablity index | ✔ | ✔ |
| 0033 | img<br>image.load_img('/content/Test/Burglary/Burglary005_x264_0.png',target_size=(64,64)) # Reading image<br>X = image.img_to_array(img) # Converting image into array<br>x = np.expand_dims (x,axis=0) # expanding Dimensions | ✔ | ✔ |
| 0034 | pred = np.argmax(model.predict(x)) # Predicting the higher probablity index<br>op = ['Fighting', 'Arrest', 'Vandalism', 'Assault', 'Stealing', 'Arson', 'Normalvideos', 'Burglary', 'Explosion', 'Robbery', 'Abuse', 'Shooting','Shoplifting', 'RoadAccide op[pred] # List indexing with output | ✔ | ✔ |
| 0035 | img = image.load_img('/content/Test/Explosion/Explosion002_x264_0.png',target_size=(64,64)) # Reading image<br>X = image.img_to_array(img) # Converting image into array | ✔ | ✔ |
| 0036 | X = np.expand_dims(x,axis=0) # expanding Dimensions<br>pred = np.argmax(model.predict(x)) # Predicting the higher probablity index | ✔ | ✔ |
| 0037 | op = ['Fighting', 'Arrest', 'Vandalism', 'Assault', 'Stealing', 'Arson', 'NormalVideos', 'Burglary', 'Explosion', 'Robbery', 'Abuse', 'Shooting', 'Shoplifting', 'RoadAccide op[pred] # List indexing with output | ✔ | ✔ |
| 0038 | img<br>image.load_img('/content/Test/Shoplifting/Shoplifting001_x264_0.png',target_size=(64,64)) # Reading image<br>X = image.img_to_array(img) # Converting image into array | ✔ | ✔ |
| 0039 | x = np.expand_dims(x,axis=0) # expanding Dimensions<br>pred = np.argmax(model.predict(x)) # Predicting the higher probablity index | ✔ | ✔ |
| 0040 | op =<br>['Fighting', 'Arrest', 'Vandalism', 'Assault', 'Stealing', 'Arson', 'NormalVideos', 'Burglary', 'Explosion', 'Robbery', 'Abuse', 'Shooting','Shoplifting', 'RoadAcci op[pred] # List indexing with output | ✔ | ✔ |
| 0041 | | | |
| 0042 | | | |
| 0043 | | | |
| 0044 | | | |

## COMMENT DETAIL

**1.1 · anyone · 20 May 2023 ·** 40 0 0 0 40 / 40 100%

(no test comments, issues or result attachments)

**2 · anyone · 20 May 2023 ·** 40 0 0 0 40 / 40 100%

(no test comments, issues or result attachments)

## ADVANTAGES & DISADVANTAGES

### Advantages:

Improved Accuracy and Precision:

- Advantage: Data analytics leverages historical project data and sophisticated algorithms, resulting in more accurate and precise project estimates.
- Impact: This leads to better cost, time, and resource predictions, reducing the likelihood of budget overruns and project delays.

Data-Driven Decision-Making:

- Advantage: Analytics-driven project estimation encourages data-based decision-making. It replaces gut feelings and intuition with concrete insights.
- Impact: Project managers can make more informed choices, enhancing overall project quality and success.

Real-time Monitoring and Adaptation:

- Advantage: Data analytics allows for real-time project monitoring, enabling rapid adaptation to changing circumstances.
- Impact: Project teams can detect issues early and take corrective actions, minimizing the impact of unforeseen challenges.

Enhanced Predictive Capability:

- Advantage: Advanced modeling and machine learning algorithms can predict potential issues or opportunities based on historical patterns.
- Impact: Project managers can proactively manage risks and seize opportunities, optimizing project outcomes.

Cost Reduction and Resource Optimization:

- Advantage: Improved accuracy in estimation reduces the chances of cost overruns and resource wastage.
- Impact: Cost savings can be significant, making the project more cost-effective.

Customization and Flexibility:

- Advantage: Data analytics solutions can be tailored to specific project requirements and industry standards.
- Impact: Customization ensures that the estimation model is aligned with the unique needs of the business project.

Scenario Analysis and Risk Mitigation:

- Advantage: Data analytics enables scenario analysis, helping project managers assess potential outcomes under different assumptions.
- Impact: By identifying and preparing for potential risks, projects become more resilient to uncertainties.

**Disadvantages:**

Data Quality Challenges:

- Disadvantage: The accuracy of estimates heavily relies on the quality of data used for analysis. Inaccurate or incomplete data can lead to unreliable predictions.
- Impact: Poor data quality may result in misleading or incorrect estimations, potentially leading to costly project mistakes.

Complex Implementation:

- Disadvantage: Implementing data analytics for project estimation can be complex and may require specialized skills.
- Impact: The complexity can result in higher initial costs and extended project timelines.

Data Privacy and Security Concerns:

- Disadvantage: Handling sensitive project data may raise concerns about data privacy and security compliance.
- Impact: Data breaches or privacy violations can have legal and reputational consequences.

Resource Intensiveness:

- Disadvantage: Data analytics can be resource-intensive in terms of hardware, software, and skilled personnel.
- Impact: Increased resource demands can elevate project costs, potentially exceeding budget limits.

Integration Challenges:

- Disadvantage: Integrating data analytics solutions with existing project management systems can be challenging and may lead to disruptions.
- Impact: Integration difficulties can slow down the adoption of analytics-driven estimation methods and affect project continuity.

Initial Setup and Learning Curve:

- Disadvantage: Setting up the data analytics system and building accurate models can be time-consuming and require a learning curve.
- Impact: Delays in project initiation and potential resistance from the team can affect project timelines.

Overreliance on Models:

- Disadvantage: Overreliance on predictive models without considering qualitative factors or expert judgment can lead to misguided decisions.
- Impact: Ignoring valuable human input and qualitative aspects may result in flawed estimations and project strategies.

Unforeseen Variables and External Factors:

- Disadvantage: Data analytics models may not account for all unforeseen variables or external factors that can affect project outcomes.

- Impact: Unexpected changes or events can disrupt project progress and accuracy of predictions.

Resistance to Change:

- Disadvantage: Introducing data analytics into the project estimation process may face resistance from team members not familiar with these techniques.
- Impact: Resistance can slow down the adoption of analytics-driven estimation methods, hindering the realization of benefits.

It's crucial to conduct a comprehensive assessment of data quality, available resources, and the organization's readiness for data analytics to maximize the benefits while mitigating the potential disadvantages in the project estimation process.

## CONCLUSION

In conclusion, the project "Estimation of Business Projects Using Data Analytics" represents a significant advancement in modern project management methodologies. By leveraging data analytics, this project aims to revolutionize the way businesses plan, execute, and monitor projects, ultimately leading to better decision-making, improved project outcomes, and increased competitiveness.

Throughout this project, we have explored the advantages and disadvantages of integrating data analytics into the project estimation process. While there are challenges, the benefits far outweigh the drawbacks, provided that the right resources and strategies are in place.

The advantages are evident in the project's ability to deliver more accurate and precise estimates, foster data-driven decision-making, enable real-time monitoring and adaptation, enhance predictive capabilities, reduce costs, offer customization and flexibility, and facilitate scenario analysis for risk mitigation. These advantages collectively contribute to the project's potential to significantly enhance project management in a business context.

However, we have also acknowledged the potential pitfalls, including data quality challenges, complexity in implementation, data privacy and security concerns, resource intensiveness, integration challenges, and the need for initial setup and learning. These disadvantages underscore the importance of careful planning, investment in training and infrastructure, and robust data governance to ensure the success of the project.

In conclusion, the "Estimation of Business Projects Using Data Analytics" project represents a transformative journey that demands a well-balanced approach to maximize its benefits while addressing its challenges. As we move forward, it is imperative to focus on data quality, security, and the development of human resources to fully realize the potential of data analytics in the estimation of business projects. With the right strategies and a commitment to data-driven decision-making, this project has the potential to reshape the landscape of project management, ultimately driving business success and growth.

## FUTURE SCOPE

The future scope for the project "Estimation of Business Projects Using Data Analytics" holds immense potential for innovation and transformation. As technology evolves, the project can explore advanced machine learning and AI integration for even more accurate and insightful predictions. It may expand into Agile project management and consider IoT and sensor data for real-time adjustments to project plans. The incorporation of blockchain can enhance data security and transparency, while seamless integration with project management software can improve user adoption. Cross-industry applications and a focus on ethical AI and responsible data use are on the horizon, and as quantum computing matures, it can be harnessed for complex modeling. The development of data-driven decision support systems, global standards, and interdisciplinary collaboration are also key elements of the future scope. With a commitment to continuous improvement and adaptation to emerging trends, this project is poised to lead the way in data-driven project estimation, offering benefits that extend across various sectors and industries.

## APPENDIX

SOURCE CODE

Github
**https://github.com/vaishwuzz/Naan-Mudhalvan/tree/main/Final%20Deliverables%20**