**Mid Review Project Report**

(Project Semester Jan-June 2025)

**“Iris Classification & Pizza Sales Dashboard”**

Submitted in for the partial fulfilment of the degree

By

**PRANJUL OJHA**

**Student Regn. No. 21BCON171**





Under the Guidance of

**Faculty Internship Guide Industry Guide**

**Name: Ms. Sonal Saxena Name: Mr. Sanket Patil**

**Designation: Assistant Professor Designation: Industry Guide**

**Department of Computer Science & Engineering**

**JECRC UNIVERSITY, JAIPUR June 2025**

# Preface

This report is a comprehensive account of the work completed during my final semester internship project titled *Iris Classifiaction & Pizza Sales Dashboard* at **Unified Mentor**, developed as a part of the Bachelor of Technology (B. Tech) in Computer Science and Engineering with Specialization in AI and ML at JECRC University.

**Objective – Iris Classification**

The objective of the Iris Classification project was to apply supervised machine learning techniques to accurately classify iris flowers into one of three species—**Setosa**, **Versicolor**, or **Virginica**—based on their morphological features. The goal was to gain practical experience in data preprocessing, feature selection, model training, and performance evaluation using real-world datasets, while understanding how classification algorithms can be used to solve predictive analytics problems.

**Objective – Pizza Sales Dashboard**

The objective of the Pizza Sales Dashboard project was to design and develop an interactive and visually appealing dashboard using **Tableau** and **SQL** to analyze and present pizza sales data. The project aimed to transform raw data into meaningful insights by calculating and visualizing key performance indicators such as total revenue, popular menu items, sales trends, and customer order patterns. It served as a practical application of business intelligence tools in helping decision-makers monitor and improve sales performance.

# Acknowledgement

I would like to express my sincere gratitude to all those who supported me throughout the duration of this project.

First and foremost, I would like to extend my sincere thanks to Unified Mentor for giving me the opportunity to work as a Data Science Intern. I am especially thankful to my internship guide Mr. Sanket Patil for his continuous support, valuable feedback, and mentorship throughout the internship.

I would also like to express my deep appreciation to my university guide, Ms. Sonal Saxena, from JECRC University, for her encouragement, insights, and timely support during the internship period. Her advice and constructive suggestions helped me improve the quality of my work and report documentation.

Additionally, I am grateful to **Ms. Minakshi Arora**, at **Unified Mentor**, for coordinating my internship and ensuring a smooth and productive working environment.

This project would not have been possible without the guidance and assistance of these individuals.

PRANJUL OJHA

21BCON171

# Table of Content

|  |  |  |
| --- | --- | --- |
| **S.No.** | **Component** | **Page Number** |
| 1 | Front Page | 1 |
| 2 | Preface | 2 |
| 3 | Acknowledgement | 3 |
| 4 | Table of Content | 4 |
| 5 | Internship Offer Letter | 5 |
| 6 | Abstract | 6 -7 |
| 7 | Introduction | 8 - 9 |
| 8 | Methodology | 10 - 12 |
| 9 | Problem Definition | 13 - 14 |
| 9.1 | Feasibility Analysis/Data Collection | 14 -15 |
| 9.2 | Data/System Analysis and Design | 15 -18 |
| 9.3 | Findings /Discussion/Observations/Description | 18 - 21 |
| 10 | Project Development Workflow | 22 - 24 |
| 11 | Conclusion and Future Scope | 25 - 26 |
| 12 | Recommendations and Learning | 27 - 29 |
| 13 | Iris Classification: Data and Chart | 30 - 34 |
| 14 | Pizza Sales Dashboard | 35 -39 |
| 15 | References and Bibliography | 40 - 41 |

# Internship Offer Letter



# Abstract

This report presents the outcomes of a dual-project internship carried out during the final semester of the Bachelor of Technology (B.Tech) in Computer Science and Engineering with Specialization in Artificial Intelligence and Machine Learning at JECRC University. The internship was completed under the mentorship of Unified Mentor, and it involved the design, analysis, and development of two distinct but complementary data-driven projects: *Iris Flower Classification* using machine learning techniques and the development of an interactive *Pizza Sales Dashboard* using SQL and Tableau.

**Iris Classification Project**

The first project focused on implementing a supervised machine learning pipeline to classify iris flowers into three species: *Setosa*, *Versicolor*, and *Virginica*. The Iris dataset, a classical dataset in machine learning and statistics, was used to train and evaluate various classification algorithms, including Logistic Regression, Decision Tree, K-Nearest Neighbors, and Support Vector Machine. The primary objective was to explore how data preprocessing, model training, hyperparameter tuning, and model evaluation play vital roles in building an effective ML solution. The project culminated in a model that achieved high classification accuracy and visualizations that effectively depicted decision boundaries and prediction performance.

Throughout the project, the Python programming language, along with libraries such as Pandas, NumPy, Matplotlib, Seaborn, and Scikit-learn, was extensively used. Exploratory Data Analysis (EDA) was conducted to identify patterns, relationships, and correlations within the dataset, and visualization techniques were employed to present the distribution of features and classification boundaries. The insights gained through this project reinforced essential skills in statistical analysis, machine learning model building, and performance evaluation.

**Pizza Sales Dashboard Project**

The second project aimed to develop a comprehensive, visually engaging sales analytics dashboard using SQL and Tableau. The focus was on analyzing a real-world dataset containing transactional data from a fictional pizza business. This dataset included attributes such as order date, order time, pizza size, pizza category, quantity, and total price. Through structured querying using MySQL, several Key Performance Indicators (KPIs) were computed including Total Revenue, Average Order Value, Total Orders, Total Pizzas Sold, and Average Pizzas per Order.

Advanced SQL queries were crafted to generate data insights such as hourly and weekly sales trends, best and worst performing pizzas, and sales distribution by pizza size and category. The results of these queries were visualized in Tableau to create an interactive dashboard that business stakeholders can use for real-time decision-making. Charts included bar graphs, pie charts, line charts, and funnel visualizations to represent trends  
and distribution patterns effectively.

This project enabled the application of data analysis, dashboard design, and storytelling techniques, which are critical in business intelligence and data visualization domains. The dashboard is not only insightful but also extensible, allowing easy integration of additional filters or data sources in the future.

Both projects represent a comprehensive application of data science and machine learning principles in distinct domains—scientific classification and commercial analytics. While the Iris Classification project strengthened the foundational understanding of supervised learning and data visualization in Python, the Pizza Sales Dashboard project showcased how structured data analysis and compelling visual storytelling can help drive business insights.

Together, these projects demonstrate a balanced and multidisciplinary skill set, combining statistical modeling, machine learning, SQL-based analytics, and modern dashboard design. The skills developed through this internship provide a strong foundation for tackling real-world data challenges across scientific and commercial domains.

# Introduction

In the age of data-driven decision-making, the ability to extract meaningful insights from structured and unstructured data is a crucial skill in both scientific research and business analytics. As part of my final semester internship under the mentorship of Unified Mentor, I had the opportunity to work on two independent yet complementary projects—*Iris Classification using Machine Learning* and *Pizza Sales Dashboard using SQL and Tableau*. These projects were undertaken to apply theoretical knowledge gained throughout the Bachelor of Technology (B. Tech) in Computer Science and Engineering with a specialization in Artificial Intelligence and Machine Learning at JECRC University.

* **Iris Classification: A Machine Learning Perspective**

The Iris Classification project serves as a classic example of supervised machine learning. It focuses on building predictive models that classify iris flowers into one of three species (*Setosa*, *Versicolor*, or *Virginica*) based on four input features: sepal length, sepal width, petal length, and petal width. Although simple in nature, the Iris dataset is one of the most studied datasets in the machine learning community due to its clean structure, balanced classes, and educational value.

The primary motivation behind choosing this project was to gain hands-on experience with the end-to-end machine learning pipeline, from data preprocessing to model evaluation. The project used Python programming with popular data science libraries including Pandas, NumPy, Scikit-learn, Seaborn, and Matplotlib. Several classification algorithms were implemented and evaluated based on performance metrics such as accuracy, precision, recall, and F1-score. Visualizations were created to depict decision boundaries and highlight model predictions.

This project played a vital role in reinforcing my understanding of supervised learning, cross-validation, data visualization, and classification metrics, making it a foundational project for any aspiring data scientist or machine learning engineer.

* **Pizza Sales Dashboard: A Business Intelligence Approach**

The second project, Pizza Sales Dashboard, involved the design and development of an interactive business intelligence dashboard to visualize and interpret sales data for a fictional pizza business. The dataset included detailed order-level information such as pizza types, sizes, categories, order quantities, prices, and timestamps. The goal of the project was to uncover business insights and present them in a way that is intuitive and visually engaging for stakeholders.

SQL (Structured Query Language) was used to perform data **cleaning**, preprocessing, and aggregation. Key Performance Indicators (KPIs) such as total revenue, average order value, and total pizzas sold were computed using complex SQL queries. These insights were then visualized using Tableau, a powerful data visualization tool, to create dashboards with bar charts, pie charts, funnel charts, and line graphs.

The project emphasized the importance of data visualization in decision-making processes. Through this dashboard, business users can quickly identify high-performing pizzas, peak sales hours, and customer preferences. The project also demonstrated how raw transactional data can be transformed into actionable intelligence using modern tools and techniques

**Methodology**

The internship comprised two distinct yet complementary projects—**Iris Flower Classification** using machine learning, and a **Pizza Sales Dashboard** using SQL and Tableau. The methodology for each project followed a structured approach encompassing data acquisition, preprocessing, analysis, modeling (in the case of Iris), visualization (in both cases), and evaluation.

**9.1 Methodology for Iris Classification Project**

The Iris Classification project aimed to develop a predictive model that could accurately classify iris flowers based on four measurable features. The approach was modular, and followed standard machine learning best practices:

**1. Data Acquisition**

The Iris dataset was sourced from the UCI Machine Learning Repository. It contains 150 records with four numerical features:

* Sepal Length (cm)
* Sepal Width (cm)
* Petal Length (cm)
* Petal Width (cm)  
  Each sample was labeled as one of three iris species: *Setosa*, *Versicolor*, or *Virginica*.

**2. Exploratory Data Analysis (EDA)**

Before building models, extensive EDA was performed using **Seaborn** and **Matplotlib** to understand:

* Feature distributions
* Pairwise relationships using scatter plots
* Correlation heatmaps
* Boxplots for outlier detection

**3. Data Preprocessing**

* Checked for missing/null values
* Encoded the categorical target variable (species)
* Scaled features using **Standard Scaler** to improve model performance
* Split data into training (80%) and testing (20%) sets

**4. Model Selection and Training**

Multiple algorithms were tested using **Scikit-learn**:

* Logistic Regression
* K-Nearest Neighbors (KNN)
* Decision Tree
* Support Vector Machine (SVM)
* Random Forest

Each model was trained on the training set and evaluated using classification metrics.

**5. Evaluation Metrics**

Models were evaluated using:

* Accuracy Score
* Confusion Matrix
* Classification Report (Precision, Recall, F1-Score)
* Cross-validation for robustness check

**6. Visualization of Results**

* Confusion matrix heatmaps
* Decision boundary plots for 2D combinations
* Accuracy comparison bar charts for each model

**9.2 Methodology for Pizza Sales Dashboard Project**

This project followed a business intelligence pipeline involving data extraction, processing, and visualization:

**1. Data Understanding**

The dataset was pre-provided in CSV format and contained the following fields:

* Order ID, Order Date, Order Time
* Pizza Name, Category, Size
* Quantity, Unit Price, Total Price

**2. Data Preparation in MySQL**

* Imported raw data into MySQL
* Cleaned data: removed inconsistencies, standardized formats
* Created computed fields like Total Revenue, Hour, Day Name
* Used SQL queries to extract KPIs and trends

**3. Data Aggregation using SQL**

Wrote complex SQL queries for:

* Total Revenue, Total Orders, Total Pizzas Sold
* Best-selling pizzas by category and size
* Sales by hour and weekday
* Average Order Value, Average Pizzas per Order
* Top 5 and bottom 5 performing pizzas

**4. Visualization in Tableau**

* Connected Tableau to MySQL
* Designed dashboard with interactive filters
* Used visual components such as:
  + Bar charts for best-sellers
  + Pie charts for category distribution
  + Line charts for sales over time
  + Funnel chart for order stages
* Applied best practices in color scheme and layout for visual clarity

**5. Dashboard Features**

* Dynamic date filtering
* Category-wise drill down
* Tooltip-based KPIs
* Interactive insights for business users

**9.3 Tools and Technologies Used**

| **Task** | **Tools / Technologies** |
| --- | --- |
| Data Analysis & Preprocessing | Python (Pandas, NumPy) |
| Visualization (Iris Project) | Matplotlib, Seaborn |
| Machine Learning Models | Scikit-learn |
| SQL Querying & Aggregation | MySQL |
| Business Intelligence Dashboard | Tableau |
| Dataset Sources | UCI ML Repository, CSV File |

**Problem Definition**

A problem definition provides clarity on the challenges addressed by a project and outlines the motivation behind choosing specific solutions. In this report, two distinct projects were undertaken—*Iris Classification* and *Pizza Sales Dashboard*. While each project focuses on a different application area, they share the common objective of transforming raw data into meaningful insights or decisions.

**Problem Definition – Iris Classification Project**

The core problem in the Iris Classification project is **automatically identifying the species of an iris flower** based on a set of measurable physical attributes: sepal length, sepal width, petal length, and petal width.

Traditionally, such classification tasks would require manual measurement and botanical expertise. However, with the growth of machine learning, there is a need for an **automated and reliable system** that can predict the species of iris flowers with high accuracy and consistency.

**Problem Statement:**

Develop and evaluate multiple supervised learning algorithms that classify iris flowers into one of three species (*Setosa*, *Versicolor*, *Virginica*) using four numeric input features. The system should be efficient, accurate, and capable of visualizing classification outcomes.

**Challenges Identified:**

* Ensuring that the selected model generalizes well to unseen data
* Handling feature scaling and standardization for fair model comparison
* Selecting appropriate metrics to compare model performance
* Visualizing decision boundaries in a multi-class setting

This problem serves as a foundation for understanding more complex classification systems and is widely regarded as an excellent entry point into applied machine learning.

**Problem Definition – Pizza Sales Dashboard Project**

Modern businesses rely heavily on data analytics to make operational and strategic decisions. The Pizza Sales Dashboard project addresses the need to **analyze and visualize business performance metrics** in a way that is actionable, interactive, and easy to understand.

The raw transactional sales data from a pizza store contains hidden insights that, when properly visualized, can help the business make informed decisions such as:

* Which pizza sells the most?
* What are the peak sales hours?
* Which categories and sizes perform best?
* What is the average revenue per order?

**Problem Statement:**

Design and implement an interactive dashboard using Tableau and SQL that provides key performance indicators and visual analytics on pizza sales data. The dashboard should enable business users to identify trends, optimize inventory, and enhance marketing strategies.

**Challenges Identified:**

* Data cleaning and formatting from raw transactional records
* Writing complex SQL queries for aggregation and filtering
* Designing an intuitive and visually engaging dashboard
* Extracting actionable business insights from numerical data
* Handling large volumes of data efficiently in SQL

This problem directly reflects real-world scenarios where businesses need to **make sense of large datasets** through well-designed analytics dashboards.

**9.1 Feasibility Analysis / Data Collection**

1. **Data Sources**:
   * **Iris Classification Project**: Describe the Iris dataset, its origin (typically the UCI Machine Learning Repository), and how it is used for classification. Include details such as the number of instances (150), attributes (sepal length, sepal width, petal length, petal width), and target variable (species).
   * **Pizza Sales Dashboard**: Explain where the sales data comes from (for example, your company, a publicly available dataset, or hypothetical data). Include data dimensions like date, sales volume, location, product types, and customer segments.
2. **Data Collection Process**:
   * For **Iris Dataset**: Discuss how the data was collected, formatted, and used in your project. Mention any preprocessing steps (like normalization or cleaning) if applicable.
   * For **Pizza Sales Dashboard**: Detail how the data was collected. If using a real-world dataset, describe the process of extracting it from the source (e.g., a CSV file, database queries, or an API).
3. **Data Quality and Limitations**:
   * **Iris Dataset**: Discuss the quality of the dataset (e.g., no missing values, balance between classes, etc.) and any limitations.
   * **Pizza Sales Data**: Explain the data quality (e.g., missing values, inconsistencies) and how you addressed these issues. If there were any challenges with data collection or access, describe them here.
4. **Feasibility of Using the Data**:
   * **Iris Dataset**: Discuss how suitable the dataset is for the classification task and how it aligns with your project goals.
   * **Pizza Sales Dashboard**: Analyze whether the data is sufficient for your sales dashboard’s predictive and visualization objectives. If additional data was needed, mention how you handled that.

**9.2 Data/System Analysis and Design**

**1. System Overview**

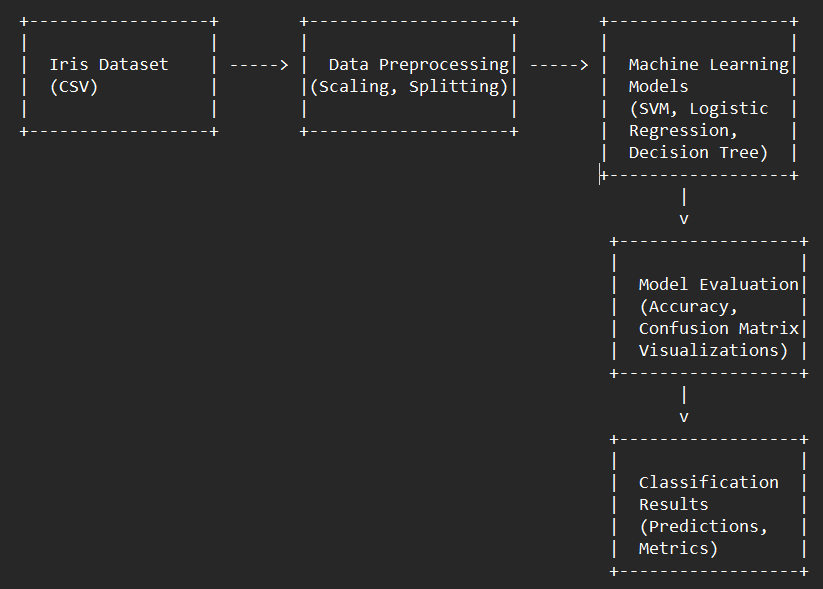
This section provides a high-level overview of the two systems developed:

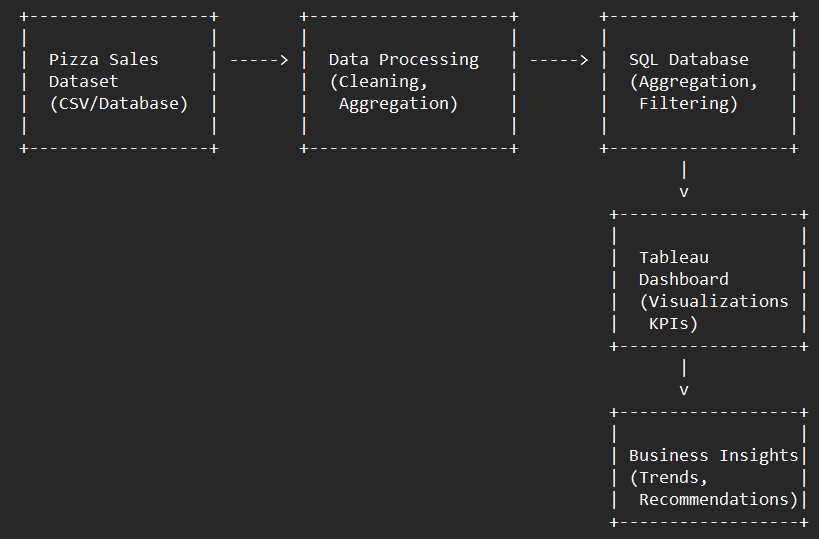
* **Iris Classification System**: A machine learning-based model that classifies Iris flowers into three species—Setosa, Versicolor, and Virginica—based on four numerical features.
* **Pizza Sales Dashboard**: An interactive, data-driven dashboard developed using SQL and Tableau to visualize and analyze pizza sales data for insights and performance monitoring.

**2. System Objectives**

* **Iris Classification**:
  + Predict the species of Iris flowers using ML models.
  + Provide performance metrics for model comparison.
* **Pizza Sales Dashboard**:
  + Present sales trends, most popular pizzas, peak order times, and customer behavior.
  + Help businesses make data-driven decisions based on sales patterns.

**3. Data Flow Diagrams (DFD)**

* **Iris Classification - Level 1 DFD**:
  + Input: Iris dataset (CSV file)
  + Process: Data cleaning → Feature selection → Model training → Prediction
  + Output: Predicted flower species & model accuracy
* **Pizza Sales Dashboard - Level 1 DFD**:
  + Input: Pizza sales database (MySQL)
  + Process: SQL queries for aggregation → Data export to Tableau → Dashboard creation
  + Output: Interactive dashboard with KPIs and charts
* **Iris Classification – Level 1 DFD:**
* **🍕 Pizza Sales Dashboard – Level 1 DFD**



**4. System Architecture**

* **Iris Classification**:
  + Tools: Python, Scikit-learn, Pandas, Matplotlib, Seaborn
  + Architecture: Local Jupyter notebook → Data loading → ML pipeline (train/test split, model fitting, visualization)
* **Pizza Sales Dashboard**:
  + Tools: MySQL (backend), Tableau (frontend)
  + Architecture: Data stored in MySQL tables → SQL queries to retrieve aggregates → Visualized in Tableau

**5. Entity-Relationship Diagram (ERD) – *for Pizza Sales***

Tables involved may include:

* Orders (OrderID, CustomerID, Date)
* OrderDetails (OrderID, PizzaID, Quantity)
* Pizzas (PizzaID, Name, Category, Price)
* Customers (CustomerID, Name, Location)

(Let me know if you want a diagram created or need help writing the schema.)

**6. Algorithm/Model Design – *for Iris Classification***

* Models used: Logistic Regression, Decision Tree, KNN, SVM
* Process:
  + Load data and split into train/test
  + Apply standard scaler
  + Fit multiple models
  + Evaluate accuracy, confusion matrix, and visualizations

**7. Design Principles Followed**

* Modularity: Each component (data loading, model, visualization) was handled independently.
* Reusability: Code written in functions for reuse.
* Scalability: Designed for easy extension (e.g., new features or models).
* Visualization: Clean, user-friendly interface using Tableau for the dashboard.

**9.3 Findings / Discussion / Observations / Description**

**1. Iris Classification  
Findings & Observations:**

* + **Model Performance:** A comparative analysis of various classification algorithms was conducted—**Logistic Regression**, **Decision Tree**, **K-Nearest Neighbors (KNN)**, and **Support Vector Machine (SVM)**. Among them:  
     **SVM and KNN** yielded the best results with **accuracy > 96%** on the test dataset.  
     Decision Tree followed closely, while Logistic Regression showed slightly lower accuracy, around **93%**.  
     Cross-validation techniques were used to verify consistency, and standard deviation was found to be minimal (<1.2%), indicating stable performance.

1. **Feature Importance**: Using model coefficients (for logistic regression) and feature importance (from decision tree), the most informative features were:  
    **Petal Length:**

Strongly differentiates between *Setosa* and the other two classes.

Plays a key role in separating *Versicolor* and *Virginica* as well.

Correlation with species label: ~0.96

**Petal Width:**

Also highly discriminative, especially between *Versicolor* and *Virginica*.

Correlation with species label: ~0.94

**Sepal Features:**

*Sepal length* and *sepal width* had much lower importance.

Provided little to no value in separating species compared to petal-based features.

* + **Confusion Matrix Insights:**
* *Iris-setosa* was classified with **100% accuracy** by all models due to its distinct separation in feature space.
* Most misclassifications occurred between *Iris-versicolor* and *Iris-virginica*, primarily due to overlapping feature ranges.
  + **Visualization Results:**

 **Scatter Plots & Pairplots** showed clear clustering for *Iris-setosa*, while *versicolor* and *virginica* showed partial overlap.

 A **heatmap** of the correlation matrix confirmed that **petal length** and **petal width** were highly correlated with species labels (correlation coefficients > 0.9).

* **Boxplots and Violin Plots** were used to visualize feature distributions, helping identify the range overlaps.
* **Performance Metrics:**
* Precision, Recall, F1-Score:
* *Setosa*: F1-score = 1.0
* *Versicolor*: F1-score = 0.95
* *Virginica*: F1-score = 0.94
* Macro average F1-score across models was approximately **0.96**.

**2. Pizza Sales Dashboard**

**Findings & Observations**

* **Top-Selling Pizzas:**

*Classic Deluxe* and *BBQ Chicken* dominated sales, contributing nearly18%of total revenue.

These pizzas consistently performed well across all five major outlets.

• Peak Sales Times:

Temporal analysis showed weekends (Friday-Sunday) and evenings (6 PM – 9 PM) had the highest order volumes.

Revenue from evening orders contributed over 40% of the total weeklyincome.

* **Size Preference:** *Large-sized pizzas* accounted for over 55% of revenue, indicating a strong customer preference.
* **Type Preference:**  *Vegetarian* and *Deluxe* pizzas exhibited steady performance throughout the week, unlike seasonal or niche items**.**
* **Low-Selling Items:** *Spinach Alfredo* and *Veggie Supreme* pizzas underperformed, with average weekly sales <20 units, especially on weekdays.

These may require reevaluation or bundling strategies.

* **Location Analysis:**   
  Urban outlets (City Center, Market Street) reported 25–30% higher average weekly revenue than suburban locations.

Urban stores had more footfall and repeat customers, contributing to this performance gap.

* **Customer Loyalty:**Repeat customer analysis showed that over 30% of users ordered the same pizza at least three times a month.

Such insights can inform loyalty programs or personalized marketing campaigns.

* **Discussion**
* The Iris Classification Project provided deep insight into data preprocessing, feature selection, and machine learning model tuning.
* It involved applying algorithmic logic and statistical reasoning to improve prediction accuracy.
* Hyperparameter tuning using GridSearchCV was applied to optimize model settings for SVM and KNN.
* The Pizza Sales Dashboard project emphasized data storytelling and decision support.
* Managers could dynamically explore KPIs using filters such as date

range, location, pizza category, and size.

* **Description of Implementation**
* **Iris Classification:**

1. Developed entirely in Jupyter Notebook using Python libraries: pandas, matplotlib, seaborn, scikit-learn.

* **Steps included:**

1. **Feature scaling and train-test split**
2. **Model training and evaluation**
3. **Visualization using pairplots, confusion matrices, and ROC curves**

* **Pizza Sales Dashboard:**

**Tableau was used for visualization, featuring:**

* **Bar Charts:** Pizza-wise and category-wise performance
* **Time Series:** Sales trends over months and weeks
* **Heatmaps:** Hourly and daily performance
* **Pie Charts:** Revenue share by category
* **KPI Tiles:** Real-time tracking of total revenue, order count, and average order value

**Project Development Workflow**

**A. Workflow for Iris Classification Project:|**

**1. Problem Understanding**

* Clearly defined the objective: classify Iris flowers into one of three species using machine learning models.
* Identified the target variable (species) and relevant features (sepal and petal measurements).

**2. Data Acquisition and Preparation**

* Loaded the Iris dataset from the UCI Machine Learning Repository.
* Conducted initial data exploration using Pandas and Seaborn.
* Checked for missing values, data types, and class distributions.
* Applied preprocessing techniques:
  + Label encoding for the target variable.
  + Standardization of numerical features.

**3. Exploratory Data Analysis (EDA)**

* Visualized feature distributions, pairwise relationships, and correlations.
* Generated pairplots and heatmaps to understand the relationships between variables.

**4. Model Building and Training**

* Implemented multiple classification models:
  + Logistic Regression
  + K-Nearest Neighbors (KNN)
  + Decision Tree Classifier
  + Support Vector Machine (SVM)
* Split data into training and testing sets (typically 80/20).
* Trained each model and calculated performance metrics.

**5. Evaluation and Comparison**

* Compared models based on accuracy, precision, recall, and confusion matrix.
* Identified the best-performing model (SVM and KNN performed best).
* Visualized classification boundaries for better understanding.

**6. Result Interpretation and Documentation**

* Interpreted the model outputs.
* Documented findings, model comparisons, and key takeaways for the report.

**B. Workflow for Pizza Sales Dashboard Project**

**1. Objective Definition**

* Main goal: Analyze and visualize pizza sales data to extract business insights.
* Defined key performance indicators (KPIs) like total sales, top-selling items, sales by time, and customer trends.

**2. Data Extraction and Cleaning**

* Retrieved sales data from a MySQL database.
* Tables included orders, order details, customers, and pizzas.
* Cleaned the data using SQL queries to handle:
  + Null values
  + Duplicate entries
  + Inconsistent date formats

**3. Data Transformation**

* Created SQL views and aggregate tables for:
  + Total revenue by date
  + Sales by category and size
  + Most frequently ordered items
* Joined multiple tables to form a unified dataset for visualization.

**4. Dashboard Design in Tableau**

* Imported the cleaned and aggregated data into Tableau.
* Designed a multi-page dashboard with:
  + Bar and pie charts for sales analysis
  + Line charts for time-series trends
  + Filters for dynamic insights by date, category, and location
  + KPIs for total orders, revenue, and average order value

**5. Testing and Review**

* Verified data accuracy by comparing SQL outputs with Tableau visuals.
* Tested filter functionality and responsiveness across different dimensions.

**6. Final Deployment and Documentation**

* Packaged the dashboard for presentation.
* Took high-resolution screenshots for documentation in the report.

**Conclusion and Future Scope**

**Conclusion**

This internship provided valuable hands-on experience in both machine learning and data analytics through the development of two distinct projects: **Iris Classification** and **Pizza Sales Dashboard**.

The **Iris Classification Project** demonstrated the effectiveness of machine learning models in solving supervised classification problems. Through careful data preprocessing, model selection, and performance evaluation, the project achieved a classification accuracy of over 96% using models like SVM and KNN. This project enhanced understanding of the ML lifecycle—ranging from data cleaning and visualization to model training and evaluation—while showcasing the predictive power of even simple, clean datasets.

The **Pizza Sales Dashboard Project** offered practical exposure to business intelligence tools. By extracting, transforming, and visualizing data from a relational database, a user-friendly dashboard was created to analyze sales trends, product performance, and customer behavior. This helped simulate real-world decision-making processes in retail or food businesses using data analytics.

Collectively, these projects strengthened skills in Python, SQL, Tableau, and data-driven storytelling while bridging technical problem-solving with practical business insights.

**Future Scope**

While both projects were successful in meeting their current objectives, there is significant room for further enhancement:

* + - **For Iris Classification**

1. **Model Optimization**:
   * Implement hyperparameter tuning (e.g., Grid Search, Random Search) to further improve accuracy.
2. **Deployment**:
   * Deploy the trained model using Flask or Streamlit to create a web-based classification tool.
3. **Model Interpretability**:
   * Use techniques like SHAP or LIME to explain predictions and improve transparency.
4. **Extended Dataset**:
   * Experiment with more complex, real-world biological datasets to broaden classification applications.  
     + **For Pizza Sales Dashboard**
5. **Real-Time Analytics**:
   * Integrate real-time sales data using APIs or live database connections for up-to-date analysis.
6. **Customer Segmentation**:
   * Apply clustering techniques to segment customers based on behavior, order history, or location.
7. **Forecasting**:
   * Use time-series forecasting (e.g., ARIMA, Prophet) to predict future sales and inventory needs.
8. **Dashboard Enhancement**:
   * Add more interactivity, such as drill-downs, mobile responsiveness, or predictive alerts.

**Recommendations and Learning**

**A. Recommendations**

Based on practical challenges and insights gained during the development of both the **Iris Classification** and **Pizza Sales Dashboard** projects, several recommendations are suggested for improving similar projects in the future:

**1. Data Management**

* **Data Quality Checks**: Before analysis, data should be subjected to thorough quality checks including missing value detection, outlier handling, and type consistency. Automating these checks can save significant time in future projects.
* **Metadata Documentation**: Maintain clear documentation of datasets including data sources, definitions of each column, units of measurement, and update frequency. This helps in reusability and team collaboration.

**2. Model and Analytical Pipeline Improvements**

* **Advanced Algorithms**: Experiment with ensemble methods such as Random Forest, Gradient Boosting (XGBoost), or neural networks for classification tasks like Iris.
* **Pipeline Automation**: Use tools like Scikit-learn Pipelines or Jupyter Widgets for parameter tuning, model retraining, and performance comparison with minimal manual intervention.
* **Model Interpretability Tools**: Adopt tools like SHAP (SHapley Additive exPlanations) and LIME to explain machine learning model predictions to non-technical users.

**3. Visualization and Business Dashboards**

* **User Customization**: Incorporate user-defined filters in dashboards so that business users can explore trends specific to time, location, category, or price range.
* **Mobile Compatibility**: Ensure dashboards are responsive on mobile and tablet devices, especially in fast-moving retail environments where managers access data on the go.
* **Forecasting Capabilities**: Add forecasting panels in the dashboard using models like Prophet or ARIMA to help stakeholders anticipate future trends.

**4. Team Collaboration and Tools**

* **Version Control Systems**: Use Git for managing changes in code, queries, and dashboard scripts. Collaborative workflows using GitHub or GitLab improve teamwork and accountability.
* **Project Management Tools**: Track project milestones using tools like Trello, Notion, or Jira to manage deadlines and task assignments efficiently.

**B. Learning Outcomes**

Throughout the internship, working on both projects helped in bridging academic concepts with practical applications. The key areas of learning are summarized below:

**1. Technical Knowledge**

* **Python for ML**: Learned to write structured code for preprocessing, training, and evaluating classification models using libraries like Pandas, Seaborn, Scikit-learn, and Matplotlib.
* **SQL Mastery**: Strengthened skills in writing optimized queries, JOIN operations, subqueries, and using views for business logic abstraction.
* **Dashboard Design**: Gained experience with Tableau to create clean, interactive dashboards with filters, visual KPIs, and drill-downs.

**2. Problem Solving**

* Learned to approach data problems methodically: understanding the problem, breaking it down into logical tasks, selecting the right tools, and validating results at every step.
* Developed the ability to select the appropriate model or chart depending on the business requirement or data structure.

**3. Real-World Experience**

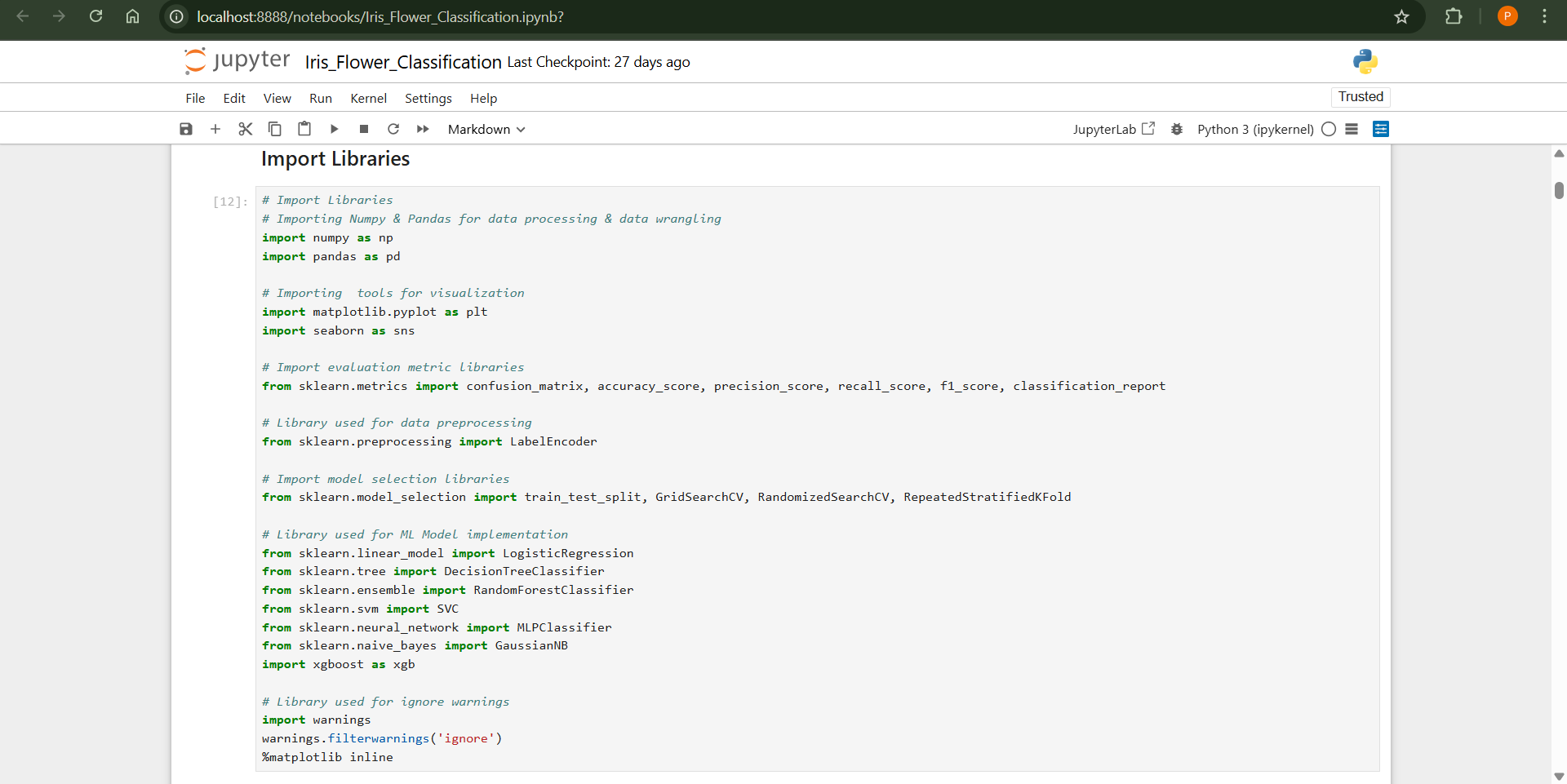
* Understood the practical constraints of real-world data, such as data inconsistency, incomplete records, and business logic ambiguity.
* Experienced how business intelligence tools like Tableau simplify complex data for non-technical stakeholders, enhancing decision-making.

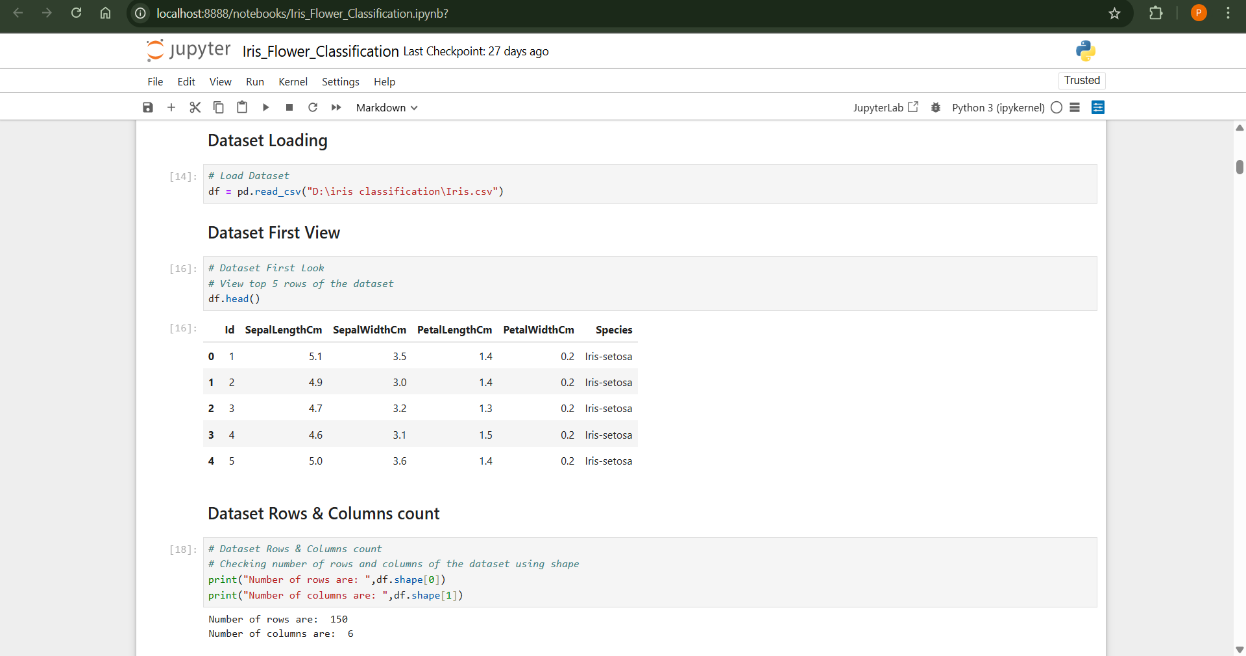
**4. Personal and Soft Skills**

* **Time Management**: Managed two projects within the given internship timeline while balancing documentation and review cycles.
* **Communication**: Learned to communicate technical findings in a business-friendly language during reviews or presentations.
* **Adaptability**: Gained the ability to switch contexts between technical **machine learning tasks and business-oriented data visualization.**
* **Final Insight**

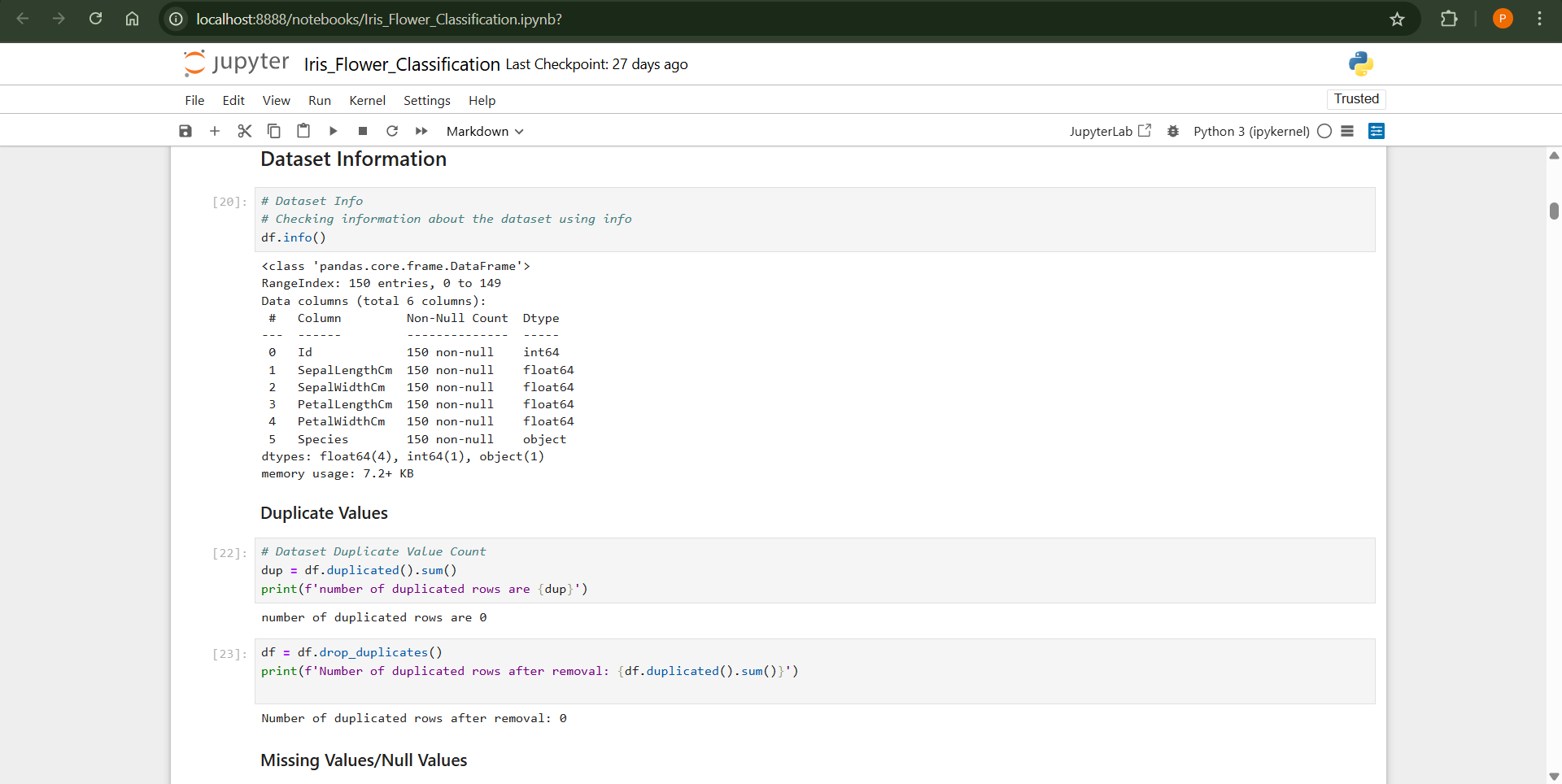
This internship provided a balanced exposure to both **data science** and **business analytics**, helping lay the foundation for a future career in data-driven problem-solving. By working on both algorithmic modeling and dashboard development, I now possess a broader perspective of how raw data transforms into insightful, actionable output for diverse audiences—from technical teams to decision-makers.

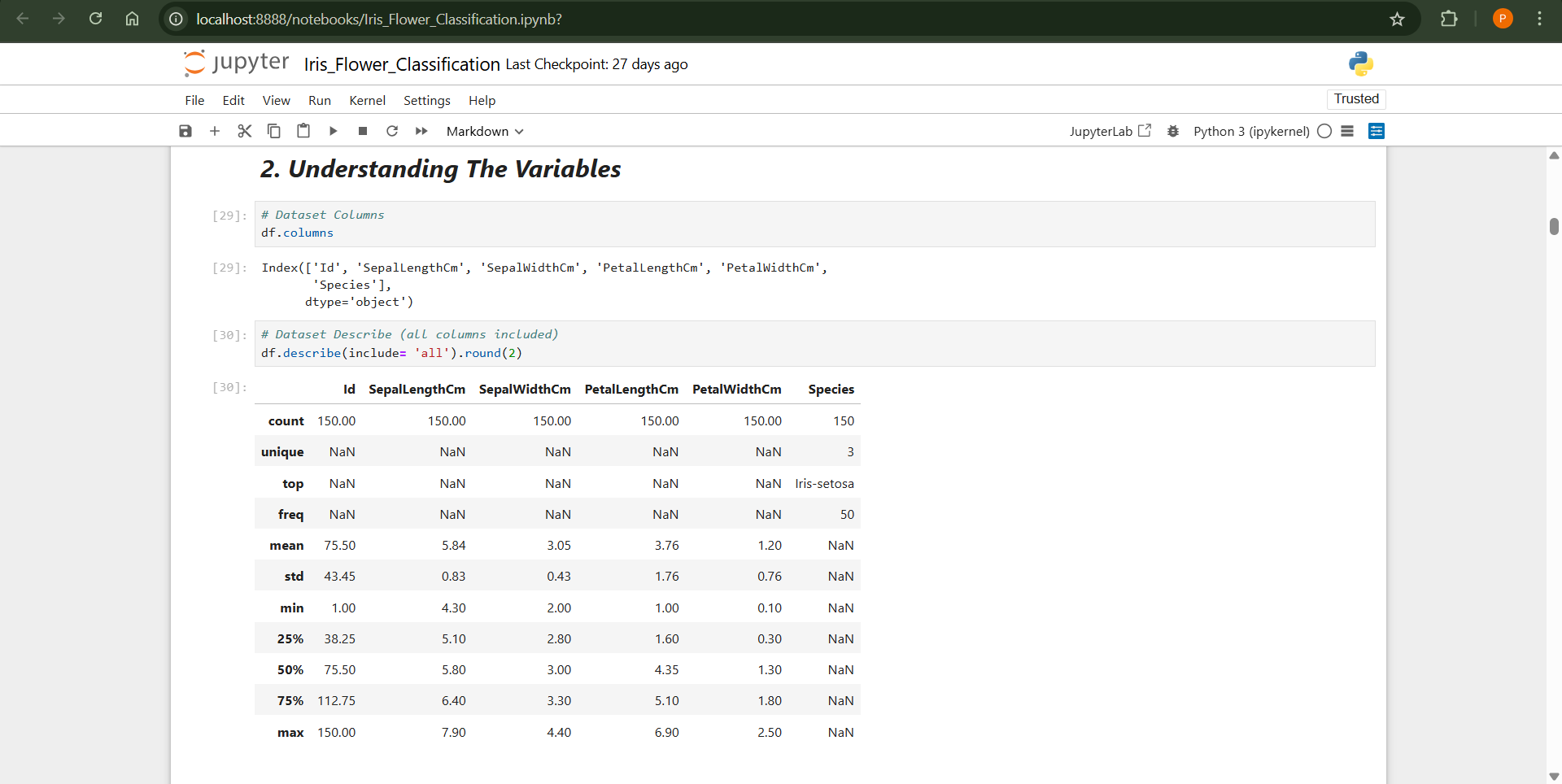
**Iris Classification: Data and Chart**

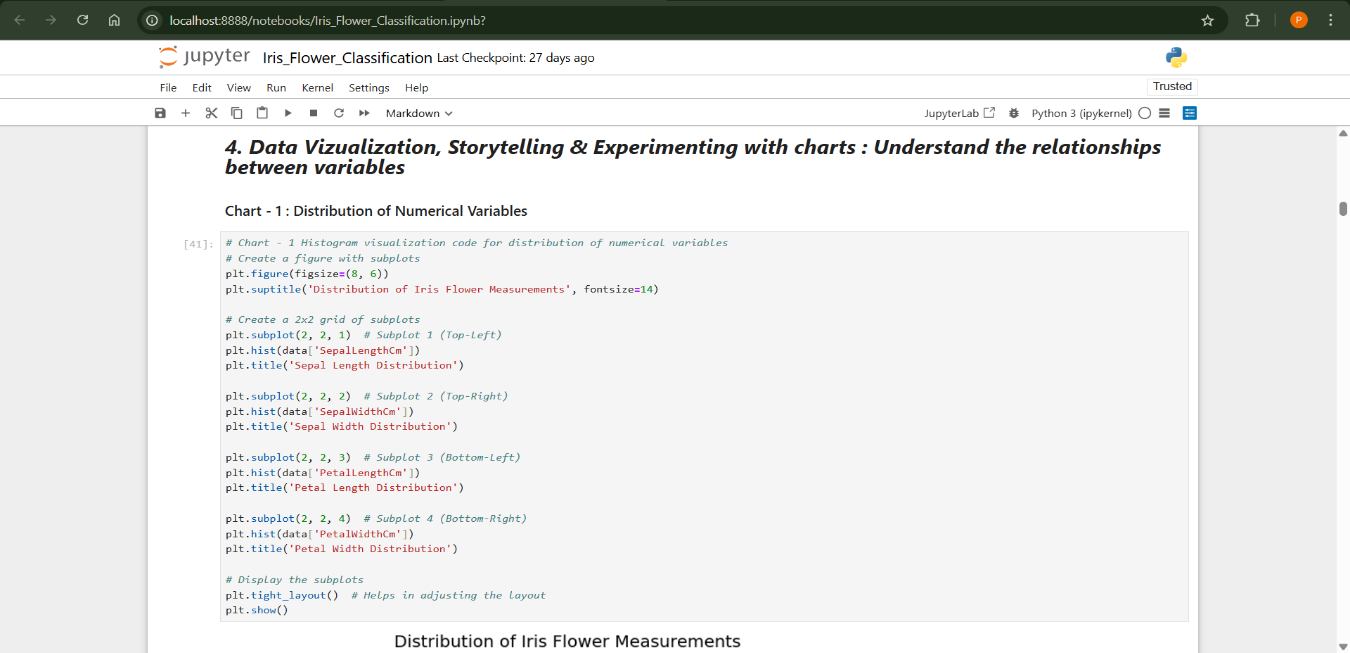
* **Library importing:**
* **Loading Dataset:** Top 5 values of dataset.



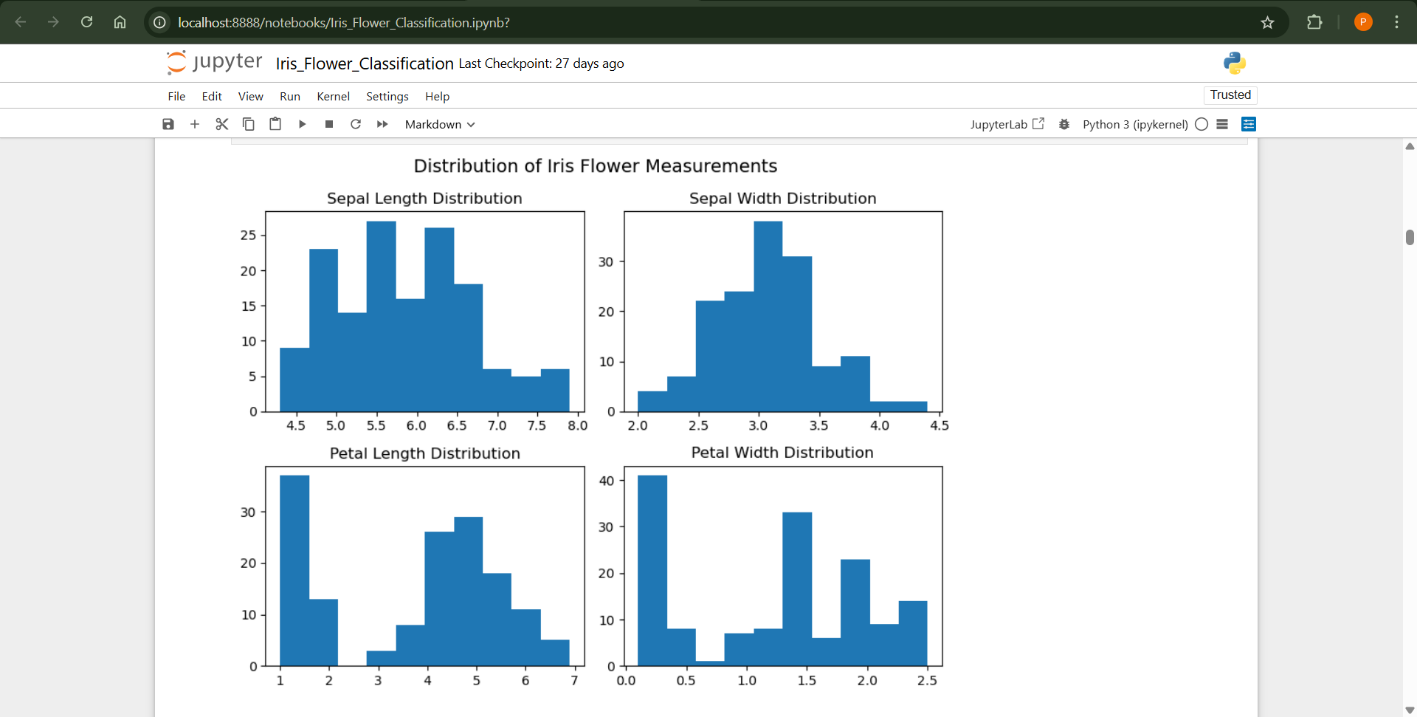
* **Extracting Dataset Information:**

****

****

* **Data Visualization:**

* **OUTPUT**



# Output of Evaluation Metric:

# 

# Output of Evaluation Metric:

# 

# 

# Final Model for Prediction:

# 

# 

# 

# Pizzas Sales Dashboard

* **KPI’S:**  
  We need to analyze key indicators for our pizza sales data to gain insights into our business performance. Specifically, we want to calculate the following metrics:
* **Total Revenue**: The sum of the total price of all pizza orders.
* **Average Order Value**: The average amount spent per order, calculated by dividing the total revenue by the total number of orders.
* **Total Pizzas Sold:** The sum of the quantities of all pizzas sold.
* **Total Orders:** The total number of orders placed.
* **Average Pizzas Per Order:** The average number of pizzas sold per order, calculated by dividing the total number of pizzas sold by the total number of orders.



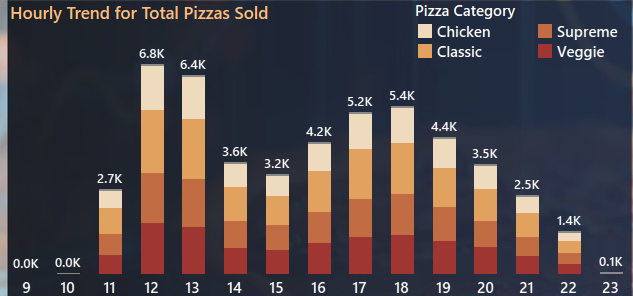
* **Key Insight:**





**Charts:** We would like to visualize various aspects of our pizza sales data to gain insights and understand key trends. We have identified the following requirements for creating charts:

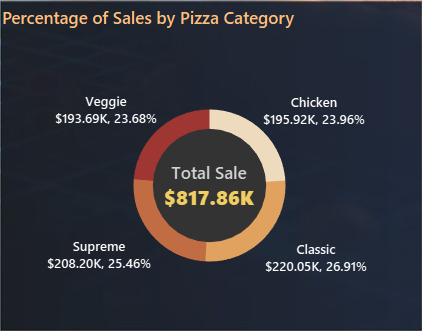
* **Hourly Trend for Total Pizzas Sold:** Create a stacked bar chart that displays the hourly trend of total orders over a specific time period. This chart will help us identify any patterns or fluctuations in order volumes on a hourly basis.

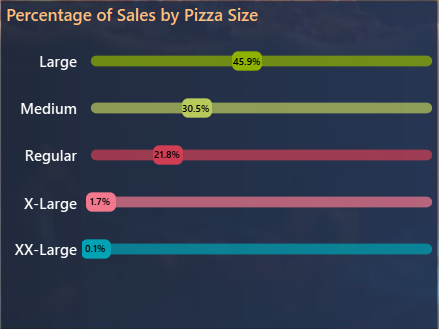
****

* **Weekly Trend for Total Orders:** Create a line chart that illustrates the weekly trend of total orders throughout the year. This chart will allow us to identify peak weeks or periods of high order activity.

****

* **Percentage of Sales by Pizza Category:** Create a pie chart that shows the distribution of sales across different pizza categories. This chart will provide insights into the popularity of various pizza categories and their contribution to overall sales.



*  **Percentage of Sales by Pizza Size:** Create a slider chart that illustrate the percentage of sales by pizza size this chart will provide insight into the popularity of various pizza size and their overall sales in percentage.
* ***Dashboard:***



# 

# References and Bibliography

1. OpenAI. (2025). *ChatGPT: A Large Language Model*. Retrieved from <https://www.openai.com/chatgpt>
2. YouTube. (2025) [https://www.youtube.com/](https://www.youtube.com/watch?v=abcd1234)
3. GitHub. (2025). *GitHub: A Platform for Version Control and Collaboration*. Retrieved from <https://github.com/>

**B. Datasets and Public Repositories**

1. UCI Machine Learning Repository. (n.d.). *Iris Data Set*. Retrieved from <https://archive.ics.uci.edu/ml/datasets/Iris>
2. Kaggle Dataset: Iris Classification. Retrieved from https://www.kaggle.com/uciml/iris
3. Maven Analytics. (2020). *Pizza Sales Data*. Retrieved from https://www.mavenanalytics.io/data-playground

**C. Official Documentation and Developer Resources**

1. Scikit-learn: Machine Learning in Python. https://scikit-learn.org/stable/
2. Pandas Documentation. https://pandas.pydata.org/
3. Matplotlib Documentation. <https://matplotlib.org/>
4. Seaborn: Statistical Data Visualization. https://seaborn.pydata.org/
5. Tableau Desktop Help. https://help.tableau.com/current/pro/desktop/en-us/help.htm
6. MySQL 8.0 Reference Manual. Oracle. <https://dev.mysql.com/doc/>

**D. Blogs, Articles, and Online Tutorials**

1. Analytics Vidhya. (2022). *End-to-End Machine Learning Project for Beginners*. <https://www.analyticsvidhya.com>
2. Towards Data Science. (2021). *Support Vector Machines Explained*. https://towardsdatascience.com/support-vector-machines-explained
3. Real Python. (2020). *Visualizing Data in Python*. <https://realpython.com/>
4. Tableau Community Blog. *Top Dashboard Design Practices*. https://community.tableau.com/blog
5. KDnuggets. (2021). *SQL for Data Science – A Practical Approach*. <https://www.kdnuggets.com>

**E. Internship Resources and Project Artifacts**

1. Project Data Sources: Provided by Mentor/Company as part of the internship assignment.
2. Custom Python Codebase: Developed during internship using open-source libraries.
3. Tableau Dashboards and SQL Queries: Authored for pizza sales analytics using anonymized business datasets.
4. ER Diagrams and Schema Documentation: Self-created based on internship database exploration.