**Republic of Tunisia**

**The Ministry of Higher Education and Scientific Research**

**ESPRIT: Private University of Engineering and Technology**

#### Data

#### Science

#### Project

#### Report

##### Academic Level : 4th Year

**Fuel Inventory Forecast System for**

**AGIL ENERGY**

#### *by*

#### Med Aziz BOUZID

Conducted within AGIL ENERGY

Une image contenant texte, logo, cheval

Description générée automatiquement

Professional Supervisor : Mr. HATEM BAHROUN

#### Academic Year: 2023 - 2024

[General Introduction](#_bookmark0) 1

1. [Business Understanding](#_bookmark1) 2
   1. [Introduction](#_bookmark2) 2
   2. [Project Overview](#_bookmark3) 2
      1. [Host Company](#_bookmark4) 2
      2. [Project Context](#_bookmark5) 2
      3. [Fuel Distribution Context](#_bookmark7) 3
      4. [Study Of The Existing](#_bookmark9) 4
      5. [Solution](#_bookmark13) 5
   3. [BOs](#_bookmark15) 6
   4. [DSOs](#_bookmark16) 6
   5. [KPIs](#_bookmark17) 6
   6. [Non-Functional Requirements](#_bookmark18) 7
      1. [Work Methodology](#_bookmark19) 7
   7. [Conclusion](#_bookmark21) 8

[Data Acquisition and Understanding](#_bookmark22) 12

* 1. [Introduction](#_bookmark23) 12
  2. [Data sets description](#_bookmark24) 12
     1. [First Dataset](#_bookmark25) 12
     2. [Second Dataset](#_bookmark32) 15
     3. [Third Dataset](#_bookmark36) 17
  3. [Data visualization](#_bookmark40) 19
     1. [First dataset](#_bookmark41) 19
     2. [Second dataset](#_bookmark45) 21
     3. [Third Dataset](#_bookmark47) 22
     4. [Conclusion](#_bookmark49) 23

[Modeling](#_bookmark52)

24

[Deployment](#_bookmark53) 25

[General Conclusion](#_bookmark54) 26

[Bibliography](#_bookmark55) 27

* 1. [AGIL ENERGY Logo](#_bookmark6) 2
  2. [TDSP Cycle](#_bookmark20) ..7
  3. [x-ray example](#_bookmark26) 12
  4. [ct-scan example](#_bookmark27) 13
  5. [round nodule example](#_bookmark28) 13
  6. [lobulated nodule example](#_bookmark29) 13
  7. [spiculated nodule example](#_bookmark30) 14
  8. [illustration of margins](#_bookmark31) 14
  9. [ct-scan1](#_bookmark33) 15
  10. [subcarinal lymph node](#_bookmark34) 16
  11. [ct-scan3](#_bookmark35) 16
  12. [ct-scan4](#_bookmark37) 17
  13. [Glandular Architecture](#_bookmark38) 18
  14. [Keratinization](#_bookmark39) 18
  15. [Union Bridges](#_bookmark42) 19
  16. [Data distribution infos](#_bookmark43) 19
  17. [Union Bridges](#_bookmark44) 20
  18. [Union Bridges](#_bookmark46) 21
  19. [Union Bridges](#_bookmark48) 22
  20. [Distribution of image counts in each category](#_bookmark50) 23
  21. [Proportion of image counts in each category](#_bookmark51) 23

In the dynamic and fast-paced world of energy distribution, effective inventory management and timely replenishment are critical to ensuring uninterrupted service at gas stations. For Agile Energy Company, forecasting fuel demand and optimizing delivery schedules are essential to prevent stockouts and minimize storage costs. With the increasing complexity of supply chains and the variability in fuel consumption patterns, there is a growing need for data-driven solutions to enhance operational efficiency.

This report delves into the development of a machine learning-based system designed to forecast fuel delivery quantities and predict the optimal delivery dates for each gas station in Agile Energy's network. By leveraging advanced data science techniques, this project aims to transform the decision-making process for fuel logistics, ensuring that fuel levels are adequately maintained while reducing the risk of overstocking or running dry.

In the first chapter, titled "**Business Understanding**," we outline the strategic importance of accurate demand forecasting and inventory optimization for gas stations. We highlight the business objectives and discuss the operational challenges that Agile Energy faces in managing its fuel supply chain.

The second chapter, "**Data Acquisition and Understanding,"** focuses on the data collection process and the complexities involved in gathering relevant historical fuel consumption and delivery data. We examine the factors influencing fuel demand and explore how these variables shape the accuracy of the predictive model.

The third chapter, "**Modeling**," details the development of the machine learning algorithms used to forecast fuel requirements and predict delivery dates. We explore various approaches, including time series analysis and regression models, and present the performance metrics that validate the model's reliability.

In the fourth chapter, "**Deployment**," we discuss the transition from model development to practical implementation. This section includes insights into the deployment process, showcasing the system's integration into Agile Energy's operations and demonstrating its real-time application in fuel management.

Finally, in the "**General Conclusion**," we provide a summary of the project, reflecting on the results achieved and the potential for future enhancements. By implementing this predictive system, Agile Energy aims to streamline its fuel delivery operations, improve inventory management, and ultimately enhance service quality across its network of gas stations.

**Chapter 1**

**Business Understanding**

## Introduction

The purpose of this chapter is to highlight the context of our work process aside with the business understanding along with the goals we aim to reach. First, we presented the host organization thereafter, we included a study of the existing and eventually, we mentioned the solution proposed for the problematic.

## Project Overview

### Host Company

The National Oil Distribution Company Agil Energy S.A. is a public company with the mission of marketing petroleum products and their derivatives under the Agil Energy label. It is one of the major public companies in Tunisia, which, through its dynamism and diverse activities, supports the national economy and ensures its continuous growth. With a turnover of 1,845 million dinars (excluding taxes) in 2020, Agil Energy S.A. plays a leading role in driving progress and excellence in the new era of Tunisia.

By expanding its activities, Agil Energy S.A. has come to occupy the leading position among companies in the sector, both in terms of sales volume and turnover, as well as the expertise of its human resources. It continually strives to strengthen this position by offering its customers the best quality products and services.

Une image contenant texte, logo, cheval

Description générée automatiquement

Figure 1.1: AGIL ENERGY Logo

### Project Context

The current landscape of fuel distribution is marked by fluctuating demand patterns, supply chain complexities, and the need for precise inventory management. For a leading company like Agil Energy S.A., which operates an extensive network of service stations across Tunisia, ensuring a steady supply of fuel to meet consumer demand is both critical and challenging.

As fuel consumption varies based on a multitude of factors—ranging from seasonal changes to economic conditions—accurately forecasting fuel demand and optimizing delivery schedules becomes essential. Mismanagement in these areas can lead to either stockouts, which disrupt operations and customer satisfaction, or overstocking, which ties up capital and increases storage costs.

In this context, the project aims to develop a robust machine learning-based system capable of predicting the quantity of fuel required at each service station and forecasting the optimal delivery dates. This system will be instrumental in enhancing the efficiency of Agil Energy S.A.’s fuel distribution operations, helping the company maintain its leading position in the market while ensuring customer satisfaction and operational excellence.

The project aligns with Agil Energy S.A.’s broader strategic objectives of leveraging advanced technology to improve decision-making processes, reduce operational risks, and support sustainable growth. By integrating data science into its operations, the company seeks to further its commitment to innovation and maintain its reputation as a pioneer in the energy sector.

### Fuel Distribution Context

Fuel distribution is a critical component of energy management that directly impacts the operational efficiency of gas stations and the satisfaction of end consumers. In Tunisia, Agil Energy S.A. plays a pivotal role in ensuring a steady supply of fuel across the country through its network of 216 service stations, 54 port stations, and 6 airport depots. The complexity of this distribution network requires a meticulous balance between maintaining adequate fuel inventory levels at each station and minimizing the costs associated with fuel storage and transportation.

The fuel distribution sector is highly dynamic and sensitive to several external factors, including economic fluctuations, geopolitical events, and environmental regulations. Seasonal demand variations, changes in fuel prices, and unforeseen disruptions such as natural disasters or supply chain bottlenecks further complicate distribution planning. Agil Energy S.A. must not only address these challenges but also anticipate shifts in fuel demand to avoid stockouts and ensure that each station remains operational.

Moreover, the transportation and storage of petroleum products carry inherent risks, including safety hazards and environmental concerns. Therefore, optimizing fuel deliveries is not just a matter of cost-efficiency but also of regulatory compliance and risk mitigation. This makes the ability to forecast fuel needs and schedule deliveries accurately a crucial aspect of the fuel distribution process.

Within this context, the project seeks to enhance the distribution efficiency of Agil Energy S.A. by developing a machine learning model that predicts fuel requirements and delivery dates for each station. By leveraging data-driven insights, the company aims to streamline its supply chain, reduce unnecessary storage and transportation costs, and maintain a reliable fuel supply that meets consumer demand in a timely and efficient manner.

### Study Of The Existing

In order to design a robust and efficient fuel demand forecasting and delivery optimization system, it is essential to first analyze the current processes and systems in place at Agil Energy S.A. for managing fuel distribution. This involves a comprehensive examination of the existing operational framework, technologies employed, and challenges faced in maintaining an uninterrupted fuel supply across its network of service stations.

### Current System Overview

Agil Energy S.A. operates under a unique business model where not every gas station bearing the "Agil" name is owned by the company. Many stations are privately owned by clients who purchase the rights to operate under the Agil brand. In such cases, Agil Energy S.A. acts as the primary fuel supplier for these stations, managing their fuel distribution needs.

Each station, whether owned by Agil or privately operated under its brand, is managed by a station manager. When fuel levels deplete, the manager is responsible for contacting Agil Energy S.A. to request a new delivery. This approach places much of the burden of fuel inventory management on the station managers, who must monitor stock levels and initiate resupply manually.

### Tools and Techniques in Use

The existing fuel management system primarily relies on station managers manually tracking fuel levels and initiating requests for replenishment. This system, while functional, lacks automation and predictive capabilities. Agil Energy S.A. uses basic inventory management software, but the communication between station managers and the distribution team is largely reactive, with little to no integration of advanced forecasting tools.

### Challenges and Limitations

Despite its functionality, the current system has several significant limitations:

1. **Manual Replenishment Requests**: The reliance on station managers to request fuel deliveries introduces delays and inefficiencies. Stockouts often occur before a new order is placed, as the system does not anticipate fuel needs proactively.
2. **Reactive Decision-Making**: The system operates on a reactive basis—fuel is ordered only when stock is low—resulting in either stockouts or excess storage, both of which are costly and inefficient.
3. **Lack of Predictive Analytics**: The absence of predictive tools means that Agil Energy S.A. cannot accurately forecast fuel needs across its network. This leads to inefficiencies in distribution, as delivery schedules and quantities are not optimized based on future demand.

### Solution

**Proposed Solution: Revolutionizing Fuel Distribution with Integrated Machine Learning Precision**

This proposal introduces an innovative solution that aims to transform the fuel distribution process for Agil Energy S.A. by integrating a sophisticated machine learning model into the existing fuel management system. The strategic deployment of this model as a dynamic platform is designed to optimize fuel deliveries, ensuring both accuracy in predicting fuel demand and timely distribution to prevent stockouts. The initial deployment will focus on the Agil network of gas stations, including both company-owned and privately operated locations.

A. Integration of Machine learning model:

The core of this proposed solution involves embedding a state-of-the-art machine learning model directly into Agil Energy's fuel management infrastructure. This integration, facilitated through a well-designed software platform, aims to enhance the current system’s predictive capabilities. By leveraging regression models, time-series forecasting, and clustering algorithms, the model promises a comprehensive approach to fuel demand forecasting and delivery optimization. The model will predict both the quantity of fuel needed and the optimal delivery dates for each station, ensuring operational efficiency and minimizing stockouts.

**B. Deployment Across Agil Network**:

The initial implementation phase will target the entire Agil network, including both the stations owned by Agil Energy S.A. and those privately operated under the Agil brand. This strategic deployment is a crucial step toward broader integration within the company's overall fuel distribution framework. By adopting this model, Agil Energy S.A. can streamline fuel deliveries across its wide network, ensuring each station is adequately supplied based on accurate demand predictions.

**C. User-Friendly Web-Based Platform**:

Complementing the machine learning model integration is the development of an intuitive web-based platform that serves as a centralized hub for fuel distribution management. This platform will allow station managers to view fuel demand forecasts, monitor current inventory levels, and track scheduled deliveries. Additionally, the platform will foster collaboration between station managers and the Agil distribution team, enabling smoother communication and ensuring that any adjustments in delivery schedules can be efficiently coordinated. This collaboration will ensure that fuel needs are met with precision and efficiency, even in fluctuating market conditions.

## Business Objectives

* 1) Define the client needs from fuel in each gas-station.
* 2) Determine the next call of the client for fuel delivery.

## Data Science Objectives

* 1) Build a predictive machine learning model in order to forecast the client needs of fuel
* 2) Build a classification machine learning model in order to predict either the client will call for fuel delivery or not

## KPIs

Key Performance Indicators (KPIs) are crucial metrics that help assess the effectiveness and success of the fuel demand forecasting and delivery optimization solution. Below are some KPIs tailored to this project:

* **Fuel Demand Forecast Accuracy (FDA)**  
  **Explanation**: This KPI measures the accuracy of the machine learning model's predictions for fuel demand at each station. A higher FDA indicates that the model is successfully forecasting the correct amount of fuel needed, reducing the risk of stockouts or overstocking.
* **Reduction in Manual Replenishment Requests (RMRR)**  
  **Explanation**: This KPI tracks the reduction in manual fuel replenishment requests made by station managers. A lower RMRR shows that the system is effectively automating the replenishment process, reducing the reliance on manual intervention.
* **Fuel Distribution Efficiency (FDE)**  
  **Explanation**: This KPI measures the overall efficiency of the fuel distribution process by comparing the amount of fuel delivered versus the amount actually needed at each station. A higher FDE indicates optimized distribution, minimizing waste and ensuring that stations receive the correct amount of fuel.

## Non-Functional Requirements

* **1) Performance:** The machine learning model must be optimized for prediction speed and computational efficiency. Techniques such as model compression and optimizations for cloud environments will be employed to reduce latency during inference. Additionally, fine-tuning the model's parameters and hyperparameters (learning rates, batch sizes, etc.) will ensure a balance between performance and accuracy, especially when processing large volumes of data across multiple stations.

* **2)Usability:** The system's interface should be intuitive and user-friendly, allowing station managers and distribution coordinators to easily access forecasts and delivery schedules. Minimal training should be required, and the platform must present data and recommendations clearly to facilitate decision-making without needing technical expertise.
* **3)Security:** The platform must ensure that sensitive operational data, including station performance and fuel demand forecasts, is protected against unauthorized access, tampering, or loss. Key security measures include:
  + **End-to-End Encryption**: Data transmitted between stations and the central system will be encrypted to ensure security during communication.
  + **Role-Based Access Control (RBAC)**: RBAC will define specific roles for different users (e.g., station managers, distribution team members) to ensure that only authorized personnel can access sensitive data.

### Work Methodology

The Team Data Science Process (TDSP) offers a structured lifecycle for data-science projects aimed at shipping as part of intelligent applications, including those deploying machine learning or artificial intelligence models for predictive analytics.

Une image contenant texte, Police, capture d’écran

Description générée automatiquement

Figure 1.6: TDSP Cycle

The TDSP lifecycle consists of five iterative stages:

* **1) Business Understanding:** This phase entails defining the objectives of the firm, identifying the main business variables that the analysis must forecast, and identifying the business challenge. This phase also defines the measures that will be used to evaluate the project’s success. Examining the data sources that are accessible and determining the type of information that is pertinent to addressing the issues that underpin the project’s objectives constitutes another crucial phase. This study will assist in determining whether further data sources or data gathering is required.
* **2)Data Acquisition and Understanding:** Being data the key ingredient of any data science project, the second stage revolves around data. It is essential to assess the current state of the data, its size and quality. In this stage, the data is ex- plored, preprocessed and cleaned. This is essential to produce a clean , high quality data set whose relationship to the target variables is understood and also to avoid propagating errors downstream and increase the chances of obtaining a reliable and accurate model.
* **3)Modeling:** After understanding and acquisition of data, feature engineering is performed on the cleaned dataset in order to generate a new, improved, dataset that facilitates model training. Feature engineering usually relies on the insights obtained from the data exploration step and on the domain expertise of the data scientist. After ensuring the dataset consists of (mostly) informative features, several models are trained and evaluated, and the best one is selected to be deployed.
* **4)Deployment:** This stage involves deploying the data pipeline and the winner model to a production or production-like environment. Model predictions can be made either in real-time or on a batch basis and this has to be decided in this stage.

## Conclusion

In this first chapter, we introduced the host company, Agil Energy S.A., and provided an overview of its significant role in Tunisia’s fuel distribution sector. We identified the core challenges in managing fuel stocks, particularly the limitations of the current system that relies on manual replenishment processes. After reviewing the existing context, we proposed a comprehensive solution that leverages machine learning to optimize fuel forecasting and distribution.

**Chapter 2**

# Data Acquisition and Understanding

## Introduction

Data acquisition and understanding form the backbone of any successful machine learning project. In this chapter, we will explore the process of gathering, analyzing, and preparing the data essential for the development of our fuel forecasting system. The quality, relevance, and structure of the data directly influence the model's performance, making this phase crucial to the overall success of the project.

## Data sets description

### Main Dataset

The main dataset used in this project consists of transactional data related to fuel deliveries for different client, across various delivery dates between 01-01-2017 and 31-12-2019. The dataset captures information about the delivery location, the type of fuel delivered, the quantity delivered, and the corresponding monetary value. Here are the columns in the dataset:

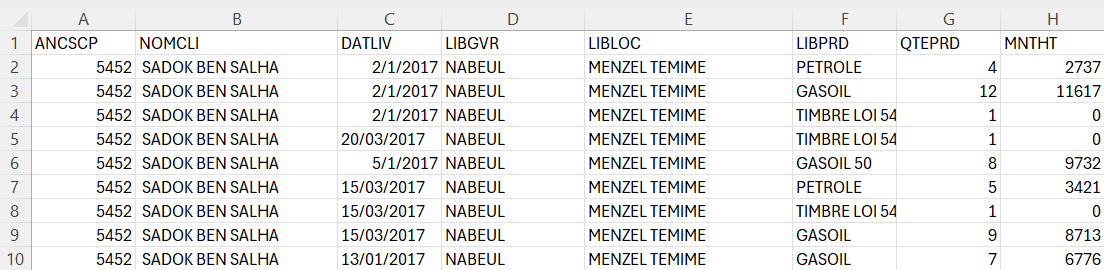


Figure 2.1: Sample of main dataset

**Features Description:**

* + **ANCSP:** Code representing the account or client number.
  + **NOMCLI:** Name of the Client
  + **DATLIV:** The delivery date of the fuel
  + **LIBGVR:** The governorate where the delivery was made

9

* + **LIBLOC:** The specific location where the fuel was delivered.
  + **QTEPRD:** The type of fuel or product delivered.
  + **MNTHT:**  The monetary value of delivery excluding taxes.

This Dataset contains 364109 rows x 8 columns.

### Tunisia Regions reference Dataset

The dataset reference table provides additional contextual information about the regions and their specific locations where fuel deliveries occur. This reference dataset is essential for understanding the geographical distribution of fuel deliveries.

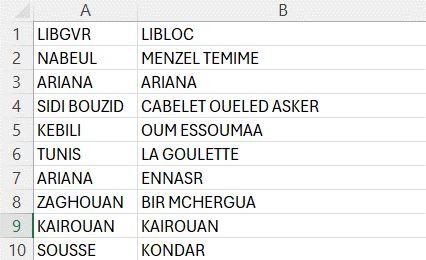


Figure 2.2: Sample of Data reference for Tunisia Region

### Client IDs Dataset:

The dataset reference table provides additional contextual information about the owners of AGIL service-station around Tunisia and their unique ID.

### Une image contenant texte, capture d’écran, Police, nombre Description générée automatiquement

Figure 2.3: Sample of Data reference for Client IDs

### 

## Data visualization

The initial data visualization phase of this lung cancer detection project focused on the distribution of CT scans across three categories: benign, malignant, and normal. The dataset composition, as depicted in the provided bar chart, revealed a majority of malig- nant cases 561 images (50%), followed by normal cases 416 images (40%) and a smaller portion of benign cases 120 images (10%). This distribution aligns with the expected higher prevalence of malignant lung cancer compared to benign cases.

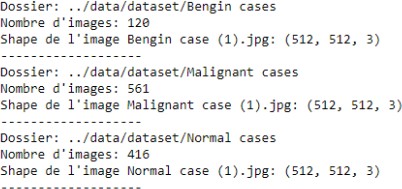


Figure 2.20: Data distribution infos

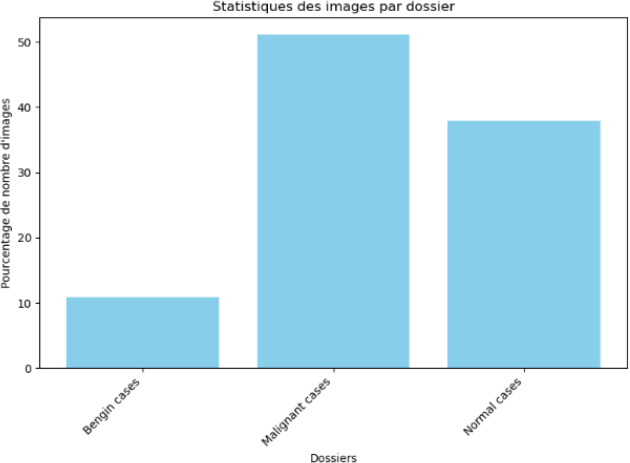
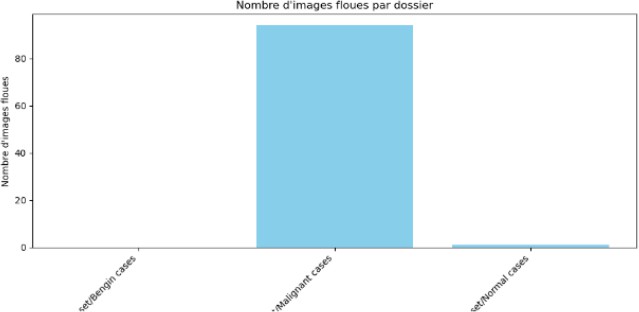


Figure 2.21: Union Bridges

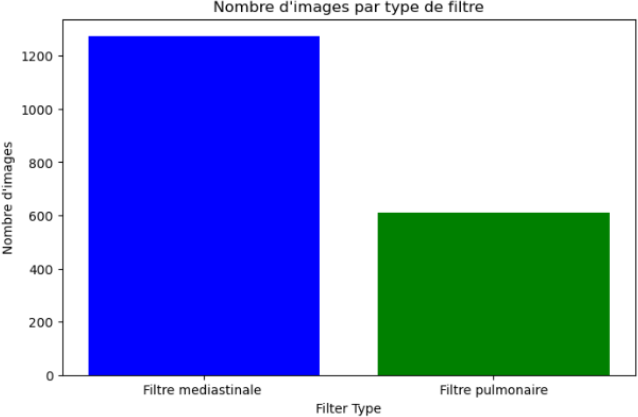
Beyond the category distribution, we also investigated the presence of blurry images within each category. The results revealed an interesting trend. Benign cases exhibited the highest image quality, with no blurry images identified (0.0%). Conversely, malig- nant cases displayed a higher prevalence of blurry images compared to normal cases, with 16.75% of their images classified as blurry compared to 0.24% for normal cases. This disparity could potentially stem from various factors, such as differences in patient positioning during scans or variations in scan quality.



### Second dataset

Figure 2.22: Union Bridges

Data Visualization of Lung Cancer CT Scans with Filters The second dataset used in this project consisted of two CT scans processed with two different filters: a mediastinal filter and a pulmonary filter. Data visualization was performed to compare the number of images generated by each filter and to assess the presence of blurry images. As shown in the bar chart, the mediastinal filter produced a significantly larger number of images (1,355) compared to the pulmonary filter (531). This difference is likely due to the inherent characteristics of each filter, with the mediastinal filter focusing on capturing details in the mediastinum, the central region of the chest, while the pulmonary filter focuses on the lungs themselves. It is important to note that no blurry images were identified in either dataset, regardless of the filter used. This finding suggests that both filters effectively preserved image quality during processing, potentially enhancing the suitability of the data for further analysis and algorithm development.



### Third Dataset

Figure 2.23: Union Bridges

The final dataset employed in this project presents a diverse collection of lung tissue images categorized into three distinct classes: lung\_n, lung\_aca , and lung\_scc . Each class comprises an equal number of images, totaling 5,000 images per class, resulting in a comprehensive dataset of 15,000 images in total. Notably, all images share a standard- ized format with a consistent 28x28 pixel dimension, ensuring uniformity and facilitating efficient processing. This standardization simplifies data handling and analysis, allowing for direct comparisons and effective feature extraction.

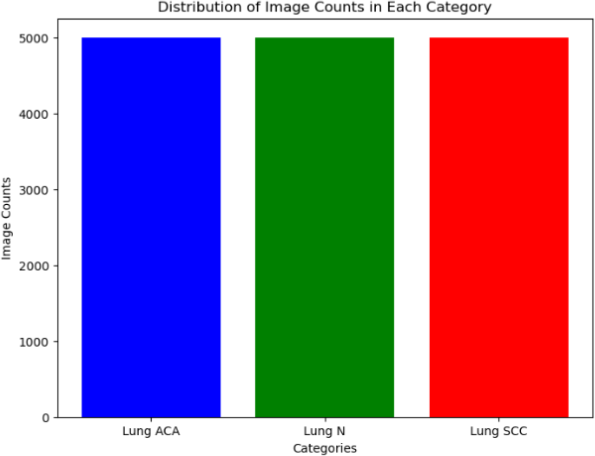


Figure 2.24: Distribution of image counts in each category

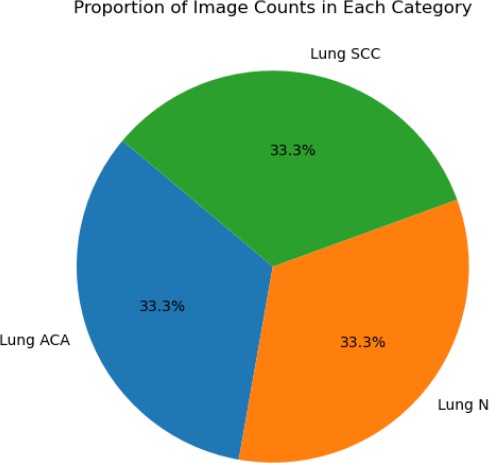


Figure 2.25: Proportion of image counts in each category

### Conclusion

In conclusion, the second chapter of the data understanding in this data science project report has provided a comprehensive exploration and analysis of the dataset. Through various techniques and methods, we have gained valuable insights into the data, enabling us to understand its characteristics, trends, and patterns

# 

# Modeling

# Deployment