Project Title: Survival Prediction for Lung Cancer Patients

Team Members:
- Andrei Tsistjakov
- Kennar Kahju
- Oliver Puusalu

Task 2: Business Understanding Report

Identifying Your Business Goals:

Background:
Lung cancer is a severe and prevalent disease with a significant impact on global health. Our project aims to leverage synthetic data to predict the survival of lung cancer patients over the next five years. Early detection and accurate survival predictions can guide medical professionals in tailoring treatment plans, improving patient outcomes, and optimizing healthcare resource allocation.

Business Goals:
1. Develop a predictive model for lung cancer survival using synthetic data.
2. Enhance the understanding of factors influencing survival rates in lung cancer patients.
3. Contribute to advancements in personalized medicine for lung cancer patients.

Business Success Criteria:
- Achieve a predictive accuracy of at least 80% in survival predictions.
- Generate insights into the most influential features affecting survival rates.

Assessing Your Situation:

Inventory of Resources:
- Access to synthetic data simulating lung cancer patient profiles.
- Expertise in data science, machine learning, and medical domain knowledge within the team.

Requirements, Assumptions, and Constraints:
- Requirement: Ensure compliance with ethical guidelines in handling patient-related data.
- Assumption: Synthetic data accurately reflects real-world lung cancer patient characteristics.
- Constraint: Limited access to certain medical databases and real patient data due to privacy and legal considerations.

Risks and Contingencies:
- Risk: Inaccuracies in synthetic data may impact model performance.
- Contingency: Conduct thorough validation and sensitivity analyses to address data quality concerns.

Terminology:

- Define medical terms and abbreviations to ensure clear communication between data scientists and healthcare professionals.

Costs and Benefits:
Costs:

Time and Effort: The primary cost may involve the time and effort invested by the team members. Developing a robust predictive model and conducting thorough analyses require dedication and expertise.

Computational Resources: While not necessarily a monetary cost, the use of computational resources, including hardware and software, could be considered a resource investment.

Benefits:

Knowledge Advancement: The project contributes to the advancement of knowledge in oncology by identifying key features influencing lung cancer survival. This intellectual benefit adds value to the scientific community.

Skill Development: Team members gain valuable experience and expertise in the intersection of data science and healthcare. The skills acquired can have long-term benefits for both individual team members and the organizations they are associated with.

Social Impact: Although not directly financial, the societal benefit of improving patient outcomes and contributing to personalized medicine for lung cancer patients is substantial. The positive impact on public health and well-being is a meaningful benefit.

Collaboration Opportunities: The project may open avenues for collaboration with healthcare professionals, researchers, and institutions. Building partnerships can lead to future collaborative projects and opportunities.

Defining Your Data-Mining Goals:

Data-Mining Goals:
1. Build a robust predictive model for lung cancer survival.
2. Identify key features influencing survival outcomes.
3. Evaluate model performance and refine as needed.

Data-Mining Success Criteria:
- Achieve a high area under the ROC curve (AUC) for the survival prediction model.
- Uncover actionable insights into the significance of various patient features on survival.
- Demonstrate the model's generalizability through rigorous testing and validation.