}

1. **PART I**

| 1. Personal Background |
| --- |
| Below is a table in which you must complete the requested information. |

| Student name | **Lucas Exequiel Carrasco Riquelme** |
| --- | --- |
| Ruth | **20.590.074-8** |
| Career | **Computer engineering** |
| Headquarters | **Antonio Varas** |

| 2. APT Project Description |
| --- |
| In the description you must briefly indicate the name of your APT project and the competencies of the graduation profile that you are going to put into practice. If performance areas are defined in your career, also mention which performance areas the project is linked to. |

| Project name | Predictive Analysis of User Behaviors in Copec Products |
| --- | --- |
| Performance Area(s) | First, data generated by products such as e-coupons and fuel cards will be analyzed to understand customer behavior and improve satisfaction. The CRISP-DM methodology will guide the analysis, using predictive models in Python to identify patterns and forecast trends, classifying customers into categories such as new, lost, declining, growing and reactivated.  Next, predictive models such as Decision Trees and Random Forest will be developed to predict future customer behavior and adjust strategies effectively.  Data visualization, with tools such as matplotlib and seaborn, will allow the results to be represented clearly, facilitating interpretation and decision making.  In addition, collaborative tools such as Google Colab and Python will be used to ensure efficient execution of the analysis.  Finally, the insights obtained will help Copec design personalized strategies and optimize resource management, improving adaptation to changes in the market and strengthening its competitive position. |
| Competencies | 1. Data Analysis: The CRISP-DM methodology will be used to analyze data from Copec digital products. This includes identifying patterns and forecasting trends using predictive models in Python, such as Decision Trees and Random Forest.  2. Development of Predictive Models: Machine learning techniques will be applied to predict the future behavior of customers, using models such as Linear Regression and Random Forest, implemented with the scikit-learn library.  3. Data Visualization: Clear and effective graphs will be created with matplotlib and seaborn to represent the results of the analysis, facilitating their understanding by stakeholders.  4. Use of Technological Tools: Google Colab will be used to collaborate in the execution of the code and Python and its libraries will be used for data analysis and modeling.  5. Optimization of Strategies and Resources: The insights obtained from predictive analysis will help Copec design more efficient marketing strategies and optimize resource management, improving adaptation to changes in consumption patterns . |

| 3. APT Project Foundation |
| --- |
| Below are different fields that you must complete with the requested information. This section seeks for you to describe your project in detail and justify its relevance and relevance. |

| Relevance of the APT project | *Every project, whether an innovation, product, service, etc., aims to respond to a situation or problem. Point out what problem your project seeks to solve and the relevance it has to the work field of your career. It also mentions the context in which this problem is situated (place, who it would impact, etc.). It is important that this problem is relevant in the context of the profession, its resolution being a real or simulated contribution to the organization or environment in which it is located. Some questions that can help you answer this section are:*  **1. Why was this topic chosen?**  The topic of predictive analysis of user behavior in Copec products was chosen due to the need to understand and anticipate customer needs in a highly competitive market such as fuel. This approach will allow Copec to optimize its marketing strategies and improve customer satisfaction, which is essential to maintain its leadership position in the sector.  **2. Why is this topic relevant to the career field?**  This topic is relevant to the labor field in data analysis and business intelligence, since it allows applying advanced methodologies and machine learning techniques to real market problems. Using predictive models to understand customer behavior and design data-based strategies is crucial in the field of data science and IT project management, strengthening the professional profile in these areas.  **3. Where is the situation to be addressed located?**  The situation is located in Chile, specifically in the context of Copec, a leading company in the fuel and associated services sector. Copec operates nationwide, so the analysis will be applied to data collected in various regions of the country, providing a comprehensive view of customer behavior in different areas.  **4. What are the main characteristics of that place?**  Copec has an extensive presence in Chile, with a network of service stations and a variety of digital products such as electronic coupons and fuel cards. The company operates in a competitive market where understanding customer preferences and behaviors is essential to adapt its offers and marketing strategies.  **5. Who is affected or impacted by the situation to be addressed?**  The situation impacts Copec customers, who use products such as electronic coupons and fuel cards. Predictive analysis will allow the identification of consumption trends and preferences of different customer segments, helping Copec to design more effective strategies to meet their needs and improve their satisfaction.  **6. What would be the value contribution (real or simulated) of the APT Project for the work and/or social context in which it would be located?**  The value contribution of the APT Project for Copec is significant in both contexts. At work, it will provide precise insights into customer behavior, allowing you to optimize marketing strategies and resource management. Socially, it will contribute to greater customer satisfaction by offering services and promotions better aligned with their preferences, improving the user experience and fostering greater brand loyalty. |
| --- | --- |
| APT Project Description | The ultimate goal is to not only accurately predict future customer behavior, but also provide Copec with valuable insights that allow it to optimize decision making in real time, improve customer retention and increase efficiency in the management of its resources. . This predictive analysis will provide a solid foundation so that Copec can anticipate the needs of its customers, quickly adapt to changes in their consumption patterns and strengthen its position in the market. |
| Relevance of the project with the graduation profile | *Justify how your APT Project is related to the graduation profile of your degree and, in particular, to the competencies of the graduation profile that you previously selected.*  **1. How is the APT Project related to the graduation profile of the degree?**  The APT Project is directly related to the graduation profile of a computer scientist by involving the application of advanced data analysis techniques in a business environment. This project requires the implementation of predictive models to understand user behavior, which is aligned with the graduation profile, which includes the ability to develop software solutions and handle large volumes of data. By applying the CRISP-DM methodology and using analysis tools such as Python, the project demonstrates how a graduate can apply their knowledge in data analysis to solve real problems in the industry, specifically in optimizing marketing strategies and improving satisfaction. of the client in Copec.  **2. In what way are the competencies you selected necessary to solve the problem to be worked on?**  The selected competencies are fundamental for the predictive analysis of user behavior in Copec products. The ability to perform advanced predictive analytics is crucial for forecasting consumer trends and customer preferences, helping to design more effective marketing strategies. The ability to implement predictive models and handle large data sets with tools such as Python ensures that the solutions developed are accurate and scalable. Additionally, project management competency ensures that analysis is carried out in a structured manner, from data collection to interpretation of results. These competencies allow us to address the problem in a comprehensive manner, offering valuable insights that improve decision-making and customer satisfaction. |
| Relationship with professional interests | *Indicate how the APT Project you propose is related to your professional interests.*  **What are your professional interests?**  My professional interests include data analytics, predictive model development, and business process optimization. I am passionate about how data can be used to make informed decisions and improve efficiency in different areas. I also have a strong interest in applying advanced machine learning and statistical analysis techniques to solve complex problems.  **What aspects of your professional interests are reflected in your APT Project?**  In my APT Project on Predictive Analysis of User Behaviors in Copec Products, these interests are clearly reflected. The project focuses on the analysis of large volumes of data to predict trends and customer behaviors, which is aligned with my passion for data analysis and the application of predictive models. I use advanced techniques such as Decision Trees and Random Forest, which allows me to work directly with machine learning and advanced statistical techniques. Additionally, the focus on optimizing marketing strategies and improving customer satisfaction also aligns with my interest in how data can transform and optimize business processes.  **Carrying out this APT Project, in what way will it contribute to your professional development?**  Completing this APT Project will significantly contribute to my professional development by offering me practical experience in the application of predictive analytics and machine learning in a real-world environment. It will allow me to develop advanced skills in data management and the implementation of predictive models, as well as data-based decision making. Additionally, working with key tools and technologies such as Python and collaborative platforms will strengthen my professional profile in the area of data science and business analytics. This experience will be invaluable to face future professional challenges and contribute to the optimization of processes in the industry. |
| Feasibility of development of the APT Project | *Briefly justify why it is possible to develop your APT project. Consider the time and materials you need to develop it, as well as the possible external factors that could hinder and facilitate its development.*  *Why do you think it is possible to develop your APT Project? To answer this question you must take into consideration:*  **1-Duration of the semester**  The project begins on August 26, 2024 and ends on December 6, 2024, covering a period of approximately 15 weeks.  **2-Hours assigned to the subject**  The project lasts 15 weeks, during which specific hours are allocated for research, development, analysis and presentation of results. It is estimated that around 20-25 hours will be dedicated per week, which adds up to an approximate total of 300-375 hours to complete the project.  **3-Required materials**  To carry out the project, the following materials will be required:   * **Data Analysis Software** : Python with libraries such as scikit-learn, pandas, numpy, matplotlib and seaborn. * **Collaboration Platforms** : Google Colab for collaborative code execution. * **Project Management Tools** : Bizagi Modeler for BPMN diagrams and tools like GitHub for version control and collaboration. * **Documentation and Reference Materials** : Documents and word processors for information management and report writing.   **4-External factors that facilitate its development**   * **Access to Tools and Resources** : The availability of platforms such as Google Colab and data analysis tools will facilitate the execution of the project. * **Collaboration and Support** : Support from tutors, mentors and colleagues can provide guidance and resolve technical doubts. * **Access to Relevant Data** : The availability of Copec user and product data will allow for exhaustive analysis and application of predictive models.   **5-External factors that hinder its development and ways in which you could solve them**   * **Limitations on Data Quality** : Data quality can be a challenge. To mitigate this, extensive cleaning and pre-processing processes will be performed to ensure the data is suitable for analysis. * **Technical Problems with Tools** : Any technical difficulties with software tools can be addressed by seeking technical support and online resources, as well as consulting with subject matter experts. * **Lack of Time** : Time management can be complicated due to multiple tasks and responsibilities. A detailed work plan will be established with milestones and intermediate deadlines to ensure that all objectives are met within the assigned period. |

1. **PART II**

| 4. Objectives |
| --- |
| In this section you must define general and specific objectives of the APT Project. It is important to clarify that the objectives must be stated clearly, concisely and without giving further explanations, that is, they must be understood on their own. It is suggested to write them using an infinitive verb, as this requires specifying specific actions. |

| General objective | The objective of this analysis is to develop a robust predictive model that allows accurately identifying and forecasting customer behavioral trends in the use of Copec products, such as electronic coupons, TCT and APR. Through detailed analysis of historical consumption data, we seek to not only categorize customers into key segments (new, lost, declining, growing, reactivated), but also understand the underlying dynamics that drive these behaviors. |
| --- | --- |
| Specific objectives | This predictive model will be designed to provide Copec with actionable insights to personalize marketing strategies, optimize resource allocation, and improve the customer experience. By identifying early patterns of change in customer behavior, you can anticipate customer churn, maximize customer lifetime value (CLV), and create more effective reactivation strategies. In addition, the analysis will allow Copec to adapt its offer and communication in real time, thus increasing customer loyalty and satisfaction, and strengthening its competitive position in the market. |

| 5. Methodology |
| --- |
| In the next section you must describe the methodology, specific to your discipline, that you will use to solve the APT project described above, including the stages and work methods. |

| Description of the Methodology |
| --- |
| *Describe how you will address the problem or situation identified above, noting the methodology that will be used to meet your objective.*  *When the project to be developed is a group project, it is necessary to incorporate the definition of the functions, tasks and responsibilities associated with each member of the team.*  Structure of predictive analysis using the CRISP-DM methodology:  **1. Understanding the business**   * **Objective:** Predict future user behaviors based on their consumption trends in products such as electronic coupons, TCT or APR. * **Key questions:**   + What are the key factors that influence customer behavior?   + What impact does each type of customer have on the business?   **2. Understanding the data**   * **Available data:** Historical sales data, purchase dates, customer identification, product type, etc. * **Relevant variables:**   + **Purchase date**   + **Purchase amount**   + **Customer identification**   + **Purchased product** * **Initial exploration:** Use **Pandas** to perform exploratory data analysis (EDA), identifying patterns, missing values, and possible anomalies.   **3. Data preparation**   * **Variable creation:**   + New Customer: Identify customers who purchased for the first time in the current period and have no purchases in the last 12 months.   + Lost Customer: Identify customers without sales in the last 3 periods and in the current period.   + Declining Customer: Calculate the average sales of the last 3 months and compare with the current sale.   + Growing Customer: Compare the current sale with the average of the last 3 months.   + Reactivated Customer: Identify customers who purchased in the current month, did not purchase in the previous 3 months, but did purchase in the same month of the previous year. * **Data Cleaning:** With **pandas** and **numpy** , handle missing values, outliers, and normalize data if necessary. * **Imputation of missing values:** If there is missing data, use **scikit-learn** 's **KNN (K-Nearest Neighbors)** to impute them based on the similarity between the data.   **4. Modeling**   * **Model selection:**   + **Decision Trees:** To classify customers based on their behavior using simple rules.   + **Random Forest Regressor:** To handle variability in data and model demand.   + **Linear or Multiple Regression:** To establish relationships between variables and predict trends in consumption. * **Training:** Fine-tune the models with a training set using **scikit-learn** . * **Validation:** Apply cross-validation with **scikit-learn** to evaluate the robustness of the models.   **5. Evaluation**   * **Metrics:** Evaluate model performance with metrics such as **R2** (coefficient of determination), **RMSE** (ratic mean square error), and **accuracy** (when applicable). * **Confusion Matrix:** If you are classifying customers, use the **Confusion Matrix** to evaluate the performance of the model. * **Visualization:** Use **matplotlib** and **seaborn** to create graphs that illustrate model performance and make results easy to interpret.   **6. Deployment**   * **Deployment:** Deploy the model to a production environment using **Google Colab** to make real-time or batch predictions. * **Collaboration and Version Control:** Use **Github** to manage code, control versions, and facilitate project collaboration. * **Monitoring:** Continue to monitor model performance and make adjustments as necessary. |

| 6. Evidence |
| --- |
| Next, describe what evidence will be evaluated in the progress report and final report of your APT project. This evidence must be agreed upon with your teacher. Evidence will be understood as the products that are developed during the project and whose purpose is to make visible or document how the work has been implemented. |

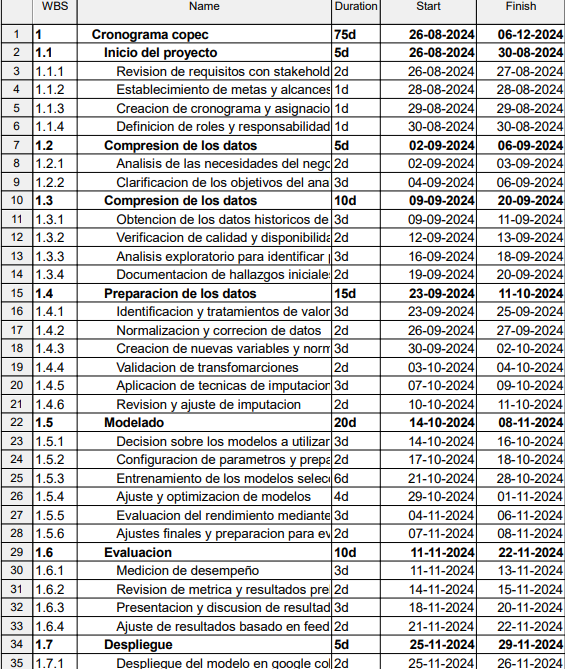
| **Type of evidence**  **(advance or end)** | **Evidence name** | **Description** | **Justification** |
| --- | --- | --- | --- |
|  |  | *Describe the evidence agreed upon with your teacher, always keeping in mind that this must account for the development of your APT Project.* |  |
| **Collaborative technical report**  **“COPEC”** | **“Copec” technical report** | A “Copec” technical report is a document that presents a detailed analysis of technical aspects of the Copec company, with data, results and conclusions relevant to decision making. | Provides a detailed and structured analysis on technical aspects of Copec. It is essential to document the progress, results and recommendations of projects or investigations, facilitating informed decision making and strategic planning based on technical data. |
| **“COPEC” word document** | **Optimization of Fuel Supply and Distribution through Predictive Analysis** | A “COPEC Word Document” is a Microsoft Word file that contains relevant and organized information about the Copec company, such as analysis, reports or proposals, in an editable textual format. | It allows you to compile and present relevant information about Copec in an editable textual format. It is useful for creating reports, proposals or any type of documentation that requires a flexible and accessible format for continuous review and editing. |
| **bizagi BPMN diagram** | **BPMN copec** | In Bizagi BPMN, "Copec" refers to a Business Process Modeling (BPMN) diagram that visualizes the business processes related to the Copec company. This diagram illustrates activities, events and workflows, making it easier to understand and improve internal processes. | Visualize and document Copec business processes using the BPMN standard. It facilitates the understanding and improvement of internal processes by graphically representing activities, events and workflows, which helps identify areas for improvement and optimization. |
| **HTML node.js prototype** | **Copec prototype** | A “Node.js Copec HTML Prototype” is a preliminary version of a web application for Copec, built using HTML for the front-end design and Node.js for the backend. This prototype allows testing and showing the basic functionality of the system before its full development. | It allows you to develop and test a preliminary version of a web application for Copec. It uses HTML for front-end design and Node.js for backend logic, making it easy to demonstrate basic functionality and get feedback before completing development. |
| **Wbs Schedule** | **Copec Predictive Analysis Schedule** | A “WBS Schedule” is a planning tool that breaks down a project into detailed tasks and subtasks, organized in a hierarchical structure known as a Work Breakdown Structure (WBS). The schedule shows the sequence and duration of each task, helping to manage and control the progress of the project | Break down the project into detailed tasks and subtasks and organize their sequence and duration. It is essential for project planning and management, as it allows you to monitor progress, allocate resources effectively, and ensure that all activities are completed on time. |

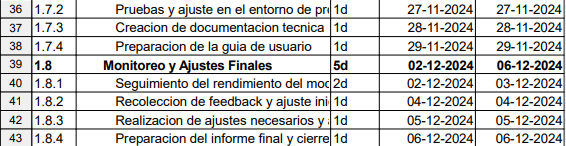
| 7. Work Plan |
| --- |
| In the following table, define the planning of your APT Project according to what is required. |

| **APT Project Work Plan** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| Competency or units of competencies | Name of Activities/Tasks | Description Activities/Tasks | Resources | Activity duration | Responsible[[1]](#footnote-0) | Observations |
| *Name the competencies or units of competencies that are related to the different activities required for the development of the activity.* | *Indicate the name of the task or activity.* | *Describe the task or activity.* | *Name the resources necessary to carry out the defined activities.* | *Write the duration of activities or tasks.* | *Write the name of the team member responsible for the activity and associated tasks.* | *Write down the difficulties or facilitators that could arise during the execution of each of the activities proposed to carry out the work plan.* |
| 1.1.1 | Review of requirements with stakeholders | The **Requirements Review with Stakeholders** consists of meeting with project stakeholders to validate and adjust the predictive analysis requirements. The goal is to ensure that all requirements are clear and aligned with business expectations. | **Requirements Documentation** : Preliminary project information.  **Communication Tools** : Platforms such as Zoom or Microsoft Teams.  **Participants** : Key stakeholders and team members.  **Technical Resources** : Computer equipment and project management tools. | 08-26-2024 until 08-27-2024 | Lucas Carrasco | **Difficulties:**  **Lack of clarity** : Difficulties in expressing needs. Solution: Detailed questions for clarification.  **Conflicts of Interest** : Disagreements between stakeholders. Solution: Effective mediation.  **Availability** : Coordination of schedules. Solution: Advance planning and flexible options.  **Facilitators:**  **Team Experience** : Knowledge of requirements management facilitates the process.  **Communication Tools** : They facilitate coordination, especially for remote work.  **Clear Documentation** : Helps in an efficient and complete review. |
| 1.1.2 | Establishment of goals and scope | Establishing **Goals and Scopes** involves clearly defining the objectives and limits of the predictive analysis project. Specific goals to be achieved are established and the scope of the project is determined to ensure that all parties involved understand what will and will not be included in the analysis. | **Project Documentation** : Information about the objectives and scope of the project.  **Planning Tools** : Project management software such as Asana or Microsoft Project.  **Participants** : Team members and relevant stakeholders.  **Technical Resources** : Computer equipment for documentation and planning. | 08-28-2024 until 08-28-2024 | Lucas Carrasco | **Difficulties:**  **Ambiguous Definition of Goals** : Goals can be vague or poorly understood. Solution: Precise writing and detailed discussion for clarity.  **Change in Scope** : Unexpected change requests may affect progress. Solution: Implement a change control process to handle modifications.  **Misalignment of Expectations** : Differences in the vision of goals between stakeholders. Solution: Meetings to align expectations and clear agreements.  **Facilitators:**  **Detailed Documentation** : Facilitates clarity in goals and scope.  **Planning Experience** : The team's experience in project planning contributes to defining goals and scopes effectively.  **Management Tools** : They facilitate the organization and monitoring of goals and scope. |
| 1.1.3 | Schedule creation and resource assignment | Creating **a Schedule and Allocation of Resources** involves developing a detailed plan that specifies the times and resources necessary for each stage of the project. A schedule with key dates is established and the necessary resources, such as personnel, tools, and budget, are assigned to ensure that the project is executed efficiently and in a timely manner. | **Project Management Software** : Tools like WBS schedule to create and manage the schedule.  **Project Data** : Information about the tasks, deadlines and resources required.  **Human Resources** : Team members with specific skills for each task.  **Material Resources** : Equipment, software and budget assigned for the project. | 08-29-2024 until 08-29-2024 | Lucas Carrasco | **Difficulties:**  **Inaccurate Estimates** : Planning can be based on incorrect estimates of time and resources. Solution: Review and adjust the schedule regularly.  **Limited Resources** : There may be restrictions on the availability of necessary resources. Solution: Prioritize tasks and look for alternatives or adjustments to the plan.  **Changes in the Project** : Modifications in the scope may affect the schedule. Solution: Implement a change management process to adjust schedule and resources.  **Facilitators:**  **Management Tools** : They facilitate the creation and monitoring of the schedule.  **Planning Expertise** : The team's experience in planning and resource allocation contributes to more accurate execution.  **Clarity in Requirements** : Having a clear understanding of the project requirements makes it easier to allocate resources appropriately and create a realistic schedule. |
| 1.1.4 | Definition of roles and responsibility | Defining **Roles and Responsibilities** involves clearly establishing the functions and duties of each team member in the project. This ensures that everyone understands their responsibilities and how they contribute to the overall success of the project, facilitating effective coordination and collaboration. | **Project Organization Chart** : Document detailing the team structure and role assignments.  **Role Descriptions** : Documents that specify the responsibilities and expectations for each role.  **Team Meetings** : Spaces to discuss and agree on roles and responsibilities.  **Project Management Tools** : Software to assign and track tasks. | 08-30-2024 until 08-30-2024 | Lucas Carrasco | **Difficulties:**  **Overlapping Responsibilities** : Poorly defined roles can lead to duplication of efforts or gaps in tasks. Solution: Define roles clearly and communicate them to the entire team.  **Role Confusion** : Team members may not understand their responsibilities. Solution: Provide training and clarification on expectations.  **Resistance to Change** : Changes in roles may be received negatively. Solution: Clearly communicate the reasons for the change and provide support during the transition.  **Facilitators:**  **Clear Documentation** : Accurate descriptions of roles and responsibilities facilitate understanding and compliance.  **Effective Communication** : Open meetings and discussions help resolve doubts and align expectations.  **Management Tools** : They facilitate the assignment and monitoring of responsibilities, ensuring that everyone is aware of their tasks. |
| 1.2.1 | Analysis of business needs | **Business Needs Analysis** involves identifying and evaluating the specific business demands and requirements that the project must address. This includes understanding current issues, improvement opportunities, and strategic objectives to ensure the predictive analytics project is aligned with business needs. | **Strategy Documents** : Information about business objectives and strategies.  **Interviews with Stakeholders** : Meetings with interested parties to collect key information.  **Analysis Tools** : Software for data analysis and requirements gathering.  **Performance Reports** : Previous reports on the current performance of the business and areas for improvement. | 09-02-2024 until 09-03-2024 | Lucas Carrasco | **Difficulties:**  **Incomplete or Ambiguous Information** : Insufficient data to fully understand business needs. Solution: Conduct multiple interviews and review additional documents to get a complete picture.  **Misalignment of Objectives** : Differences in the vision of objectives between stakeholders. Solution: Facilitate goal alignment and clarification sessions.  **Resistance to Change** : Resistance to proposed changes based on the analysis. Solution: Communicate clearly about the benefits of the project and how it will address identified needs.  **Facilitators:**  **Access to Stakeholders** : Effective meetings and interviews with stakeholders provide valuable information.  **Comprehensive Documentation** : Having access to strategic documents helps to better understand business needs.  **Analysis Tools** : Facilitate data evaluation and identification of key needs. |
| 1.2.2 | Clarifying the goals of predictive analytics | Clarifying **Predictive Analytics Goals** involves defining and detailing the specific goals that predictive analytics should achieve. This includes establishing what outcomes are expected, how they will be measured, and how they will contribute to business decision-making. | **Requirements Documents** : Previous information about the project objectives.  **Definition Meetings** : Spaces to discuss and agree on the objectives of the analysis with stakeholders.  **Planning Tools** : Software to document and manage objectives, such as Google Docs or Microsoft Word.  **Historical Data** : Previous data that will serve to define and validate the objectives of the analysis. | 09-04-2024 until 09-06-2024 | Lucas Carrasco | **Difficulties:**  **Vague or Incomplete Objectives** : Lack of clarity in objectives can hinder effective implementation. Solution: Detail the objectives and ensure their alignment with business expectations.  **Disagreements between Stakeholders** : Differences in the expectations of the objectives. Solution: Facilitate discussions to align expectations and define agreed-upon objectives.  **Changes in Requirements** : Modifications in objectives as the project progresses. Solution: Implement a change management process to adjust goals as necessary.  **Facilitators:**  **Effective Meetings** : Discussing the objectives with stakeholders facilitates a clear and agreed definition.  **Clear Documentation** : Having a detailed record of objectives helps maintain focus during the project.  **Access to Historical Data** : Using previous data helps define realistic and measurable objectives. |
| 1.3.1 | Obtaining historical sales and customer behavior data | Obtaining **Historical Sales and Customer Behavior Data** involves collecting relevant data about past sales and customer behavior patterns. This data is essential to perform accurate predictive analysis, allowing trends and patterns to be identified that will inform future forecasts. | **Databases** : Access to data management systems that store sales and customer information.  **Data Extraction Tools** : Software to query and extract data, such as SQL or ETL tools.  **Data Documentation** : Guides or manuals on the structure and origin of the data.  **Human Resources** : Personnel with skills in data extraction and analysis. | 09-09-2024 until 09-11-2024 | Lucas Carrasco | **Difficulties:**  **Data Access** : Restrictions on access to databases or incomplete data. Solution: Coordinate with the IT team to ensure data access and integrity.  **Data Quality** : Data that is outdated or has errors. Solution: Implement data cleaning and validation processes to ensure its quality.  **Data Integration** : Difficulties in combining data from different sources. Solution: Use integration tools and standardize data formats.  **Facilitators:**  **Access to Data Systems** : Facilitates obtaining data necessary for analysis.  **Extraction Tools** : Allow efficient data recovery.  **Complete Documentation** : Helps understand the structure and origin of the data, facilitating its use. |
| 1.3.2 | Data quality and availability verification | **Data Quality and Availability Verification** consists of evaluating and ensuring that the data obtained meets the quality criteria and is available for analysis. This includes checking the accuracy, completeness, consistency and timeliness of the data, as well as its accessibility and suitability for the project objectives. | **Data Analysis Tools** : Software such as Python (Pandas) or R to perform quality analysis.  **Data Quality Checklist** : Document with specific criteria to evaluate the quality of the data.  **Data Documentation** : Information about the structure and origin of the data.  **IT Teams** : Technical support to resolve data access or integrity issues. | 09-12-2024 until 09-13-2024 | Lucas Carrasco | **Difficulties:**  **Incomplete or Incorrect Data** : Missing or incorrect data can affect the analysis. Solution: Implement data cleaning and validation techniques.  **Restricted Data Access** : Problems accessing necessary data. Solution: Work with the IT team to ensure proper access.  **Data Inconsistencies** : Differences in data formats or structures. Solution: Establish standards and perform necessary transformations to standardize data.  **Facilitators:**  **Analysis Tools** : Facilitate accurate assessment of data quality.  **Complete Documentation** : Helps understand the integrity and context of the data.  **Technical Support** : Allows you to quickly resolve data access or quality problems. |
| 1.3.3 | Exploratory analysis to identify patterns and problems | Analysis **to Identify Patterns and Problems** involves examining collected data to discover patterns, trends, and potential problems. This preliminary analysis allows you to gain a general understanding of the data, detect anomalies, and lay the foundation for more detailed predictive analysis. | **Data Analysis Tools** : Software such as Python (pandas, matplotlib, seaborn) or R to perform visualizations and statistical analysis.  **Data Documentation** : Information about the structure and origin of the data to understand the context.  **Initial Reports** : Preliminary results of the analyzes for review and discussion.  **Data Teams** : Personnel with skills in data analysis and statistics. | 09-16-2024 until 09-18-2024 | Lucas Carrasco | **Difficulties:**  **Messy Data** : Disorganized or poorly structured data can make analysis difficult. Solution: Perform prior cleaning and preparation of the data.  **Complex Pattern Detection** : Identifying patterns in large volumes of data can be challenging. Solution: Apply visualization techniques and advanced statistical analysis.  **Interpretation Problems** : Difficulty in interpreting preliminary results. Solution: Consult with experts or perform complementary analyzes to validate findings.  **Facilitators:**  **Analysis Tools** : They facilitate the exploration and visualization of data.  **Detailed Documentation** : Provides context that helps interpret analysis results.  **Analysis Experience** : Skills in data analysis and statistics allow for effective identification of patterns and problems. |
| 1.3.4 | Documentation of initial findings | Documentation **of Initial Findings** involves recording and detailing the discoveries and observations obtained during exploratory data analysis. This process includes the creation of reports or summaries that describe the patterns identified, problems detected, and any relevant findings that may influence subsequent analysis. | **Documentation Tools** : Software such as Microsoft Word, Google Docs, or reporting tools such as Jupyter Notebooks to record findings.  **Report Templates** : Predefined structures to organize findings clearly.  **Data and Visualizations** : Results and graphs generated during the exploratory analysis.  **Work Team** : Personnel in charge of writing and reviewing the documentation. | 09-19-2024 until 09-20-2024 | Lucas Carrasco | **Difficulties:**  **Clarity in Documentation** : Challenges in clearly expressing findings. Solution: Use templates and examples to structure information coherently.  **Data Inconsistencies** : Contradictory findings can make documentation difficult. Solution: Review and validate data and results before documentation.  **Limited Time** : Pressure to document findings quickly. Solution: Establish appropriate deadlines and priorities to ensure thorough documentation.  **Facilitators:**  **Documentation Tools** : They facilitate the organization and clear presentation of the findings.  **Report Templates** : They provide a structure that helps organize information effectively.  **Data and Visualizations** : Help support findings with concrete evidence, making it easier to communicate results. |
| 1.4.1 | Identification and treatments of missing values and anomalies | Identification **and Treatment of Missing Values and Anomalies** consists of detecting and handling incomplete or unusual data in the data set. This includes identifying missing values, anomalies or errors in the data and applying appropriate methods to correct or impute these problems, thus ensuring the integrity and quality of the analysis. | **Data Analysis Tools** : Software such as Python (pandas, numpy) or R to detect and deal with missing data and anomalies.  **Imputation Methods** : Techniques to replace missing values, such as imputation by the mean, median, or advanced methods such as KNN Imputer.  **Procedures Documentation** : Guides on how to deal with different types of anomalies and missing values.  **Data Team** : Personnel with skills in data analysis and cleaning. | 09-23-2024 until 09-25-2024 | Lucas Carrasco | **Difficulties:**  **Identification of Anomalies** : Difficulty detecting outliers or errors in large volumes of data. Solution: Use visualization tools and statistical techniques to identify anomalies.  **Choosing the Imputation Method** : Deciding on the best method for handling missing values can be complex. Solution: Evaluate different imputation methods and select the most appropriate one based on context and impact on data.  **Impact on Analysis** : Handling of missing values and anomalies can affect data quality. Solution: Implement validation strategies to ensure that the treatment does not introduce bias.  **Facilitators:**  **Analysis Tools** : They facilitate the detection and treatment of problems in the data.  **Established Methods** : There are proven methods and techniques for handling missing values and anomalies that can be applied.  **Documentation and Procedures** : They help standardize data processing, ensuring consistency and precision in the process. |
| 1.4.2 | Data normalization and correction | **Data Normalization and Correction** involves adjusting the data so that it is in a uniform format and correcting any errors to ensure its consistency and accuracy. Normalization adjusts the scales and formats of data to make it easier to compare and analyze, while correction fixes errors or inconsistencies in the data. | **Data Analysis Tools** : Software such as Python (pandas, numpy) or R to perform data normalization and correction.  **Normalization Methods** : Techniques such as data scaling (min-max, standardization) to standardize the scales.  **Data Correction Guides** : Documentation on procedures for correcting common errors in data.  **Data Documentation** : Information on the required data format and rules. | 09-26-2024 until 09-27-2024 | Lucas Carrasco | **Difficulties:**  **Format Inconsistencies** : Different data formats can be difficult to standardize. Solution: Implement data conversion and standardization processes.  **Complex Errors** : Correcting errors in data can be complicated if the errors are varied or not well documented. Solution: Establish clear procedures to identify and correct errors.  **Loss of Information** : Normalization can lead to a loss of information or the introduction of bias. Solution: Evaluate the impact of normalization and make adjustments as necessary.  **Facilitators:**  **Analysis Tools** : They facilitate the application of normalization and correction techniques.  **Normalization Methods** : Provide standardized approaches to fit data uniformly.  **Documentation and Procedures** : They offer clear guidelines on how to perform data normalization and correction, ensuring a consistent and accurate process. |
| 1.4.3 | Creation of new variables and normalization | Creation **of New Variables and Normalization** consists of generating new variables from existing data to improve the analysis and then normalizing these new variables to ensure consistency in the model. Creating new variables can include deriving additional features that provide more information about the data, while normalization ensures that all variables are in a comparable format and scale. | **Data Analysis Tools** : Software such as Python (pandas, numpy) or R for the creation and normalization of variables.  **Variable Creation Methods** : Techniques for deriving new characteristics (eg, categorical to numerical variables, combinations of variables).  **Normalization Methods** : Techniques such as scaling (min-max, standardization) to adjust the new variables.  **Variable Documentation** : Information about the logic behind the creation of new variables and the normalization rules applied. | 09-30-2024 until 10-02-2024 | Lucas Carrasco | **Difficulties:**  **Defining Relevant Variables** : Determining which new variables will be useful can be complicated. Solution: Perform a prior analysis of the data and consult with experts to identify significant characteristics.  **Consistency in Normalization** : Ensure that new variables are correctly normalized to avoid bias. Solution: Apply standardized normalization methods and review the results.  **Impact on the Model** : New variables may not always improve the model and may introduce additional complexity. Solution: Evaluate the impact of each new variable through testing and validation.  **Facilitators:**  **Analysis Tools** : They facilitate the creation and normalization of new variables efficiently.  **Standardized Methods** : They provide systematic approaches to creating and normalizing variables, ensuring consistency and precision.  **Documentation and Procedures** : They help maintain a clear record of the new variables and the normalization process, facilitating replication and review. |
| 1.4.4 | Transformation validation | **Transformation Validation** involves verifying that transformations applied to the data, such as creating new variables or normalizing, have been performed correctly and are suitable for analysis. This includes checking that transformations have not introduced errors, distorted data, or affected data integrity. | **Data Analysis Tools** : Software such as Python (pandas, numpy) or R to review and validate the applied transformations.  **Validation Methods** : Techniques to check the accuracy and impact of transformations (eg, distribution analysis, comparison of statistics before and after the transformation).  **Documentation of Transformations** : Detailed record of the transformations carried out and the validation criteria used.  **Data Team** : Personnel specialized in data analysis to review and confirm the validity of the transformations. | 03-10-2024 until 04-10-2024 | Lucas Carrasco | **Difficulties:**  **Error Identification** : Detecting problems in transformations can be difficult, especially if they are subtle. Solution: Implement rigorous testing and review the data in detail to identify discrepancies.  **Impact on Analysis** : Transformations can have unexpected effects on the analysis. Solution: Perform sensitivity analysis to evaluate the impact of transformations on results.  **Complexity of Transformations** : Some transformations can be complex and difficult to validate. Solution: Use standard validation tools and techniques and consult with experts if necessary.  **Facilitators:**  **Analysis Tools** : They facilitate the review and validation of transformations efficiently.  **Standardized Validation Methods** : They offer clear procedures to evaluate the precision and effectiveness of transformations.  **Detailed Documentation** : Helps keep track of the transformations carried out and ensure that all steps are properly reviewed. |
| 1.4.5 | Application of imputation techniques using KNN | The **Application of Imputation Techniques using KNN** consists of using the K-Nearest Neighbors (KNN) algorithm to replace missing values in the data set. This imputation method estimates missing values based on the values of the closest instances in the feature space, allowing consistency in the data to be maintained. | **Data Analysis Tools** : Software like Python with the scikit-learn library to implement the KNN algorithm.  **Data Set** : Data with missing values that need to be imputed.  **Technical Documentation** : Guides and documentation on the use of KNN for data imputation.  **Data Team** : Personnel with skills in data imputation and modeling techniques. | 07-10-2024 until 09-10-2024 | Lucas Carrasco | **Difficulties:**  **Selection of Number of Neighbors (k)** : Choosing the optimal number of neighbors can be challenging and affect the quality of imputation. Solution: Perform tests with different values of k and evaluate the performance to select the most suitable one.  **Performance on Big Data** : KNN can be computationally intensive with large volumes of data. Solution: Optimize the algorithm or consider alternative methods if performance is an issue.  **Impact of Imputation** : Imputation can introduce bias if not handled correctly. Solution: Evaluate the accuracy of the imputation and compare the results before and after the process.  **Facilitators:**  **Analysis Tools** : They facilitate the implementation and adjustment of the KNN algorithm for imputation.  **KNN Method** : Provides an effective approach for imputation based on similarities between data.  **Technical Documentation** : Offers clear guidelines to apply the method appropriately and efficiently. |
| 1.4.6 | Allocation review and adjustments | **Imputation Review and Adjustments** involves evaluating and adjusting the data imputation process to ensure that missing values have been replaced appropriately and that the resulting data is accurate and useful for analysis. This activity involves reviewing the results of the imputation performed, detecting any errors or anomalies, and adjusting the process as necessary to improve data quality. | **Data Analysis Tools** : Software such as Python (scikit-learn, pandas) to review and adjust data imputation.  **Data Set** : Data with imputations made that need to be reviewed.  **Evaluation Methods** : Techniques to evaluate the quality of imputation, such as comparison of descriptive statistics before and after imputation.  **Imputation Documentation** : Record of the methods and parameters used during imputation to facilitate review and adjustments. | 10-10-2024 until 10-11-2024 | Lucas Carrasco | **Difficulties:**  **Error Identification** : Detecting errors or problems in imputed data can be complicated. Solution: Perform detailed analysis and comparisons with original data to identify and correct errors.  **Complex Adjustments** : Adjusting the imputation process can be technical and require fine adjustments. Solution: Use advanced analysis tools and consult with data experts if necessary.  **Impact on Analysis** : Settings may affect analysis results in unexpected ways. Solution: Evaluate the impact of adjustments to the model and adjust based on the results.  **Facilitators:**  **Analysis Tools** : They facilitate the review and adjustment of imputed data efficiently.  **Evaluation Methods** : They provide systematic ways to verify the quality of imputation and ensure that the data is correct.  **Detailed Documentation** : Helps maintain a clear record of the processes and adjustments made, facilitating review and correction. |
| 1.5.1 | Decision on the models to use | The **Decision on the Models to Use** involves selecting the most appropriate algorithms and modeling techniques for predictive analysis based on the characteristics of the data set and the objectives of the project. This activity requires evaluating different models, such as Decision Trees, Random Forest, Linear Regression, among others, and choosing those that best adapt to the needs of the analysis and provide the most accurate results. | **Modeling Tools** : Software such as Python (scikit-learn) to implement and evaluate different predictive analysis models.  **Data Set** : Historical and transformed data that will be used to train and evaluate the models.  **Documentation and Literature** : Guides and articles on the available models, their characteristics and their applications.  **Data Team** : Personnel with experience in predictive modeling and algorithm selection. | 10-14-2024 until 10-16-2024 | Lucas Carrasco | **Difficulties:**  **Model Evaluation** : Determining which model is the most suitable can be complicated due to the variety of models and parameters. Solution: Use cross-validation and benchmarking techniques to evaluate the performance of each model.  **Computational Requirements** : Some models can be intensive in resources and processing time. Solution: Optimize the code and consider less complex models if resources are limited.  **Interpretation of Results** : Interpretation of results from different models can be difficult. Solution: Use visualization and analysis tools to facilitate understanding of the results.  **Facilitators:**  **Modeling Tools** : They facilitate the implementation and testing of different models efficiently.  **Documentation and Literature** : Provides valuable information about the models and their applications, helping to make informed decisions.  **Team Expertise** : Staff with experience in predictive modeling can guide the selection of the most suitable models and optimize the process. |
| 1.5.2 | Setting parameters and preparing for development | Configuration **and Preparation for Development** involves adjusting the parameters of the selected models and preparing the development environment for deployment. This task includes the configuration of hyperparameters in the predictive models, the preparation of the programming environment and the definition of the tools and libraries necessary to carry out the analysis and development of the project. | **Development Tools** : Software such as Python and its libraries (scikit-learn, pandas, numpy) for model adjustment and configuration.  **Development Environment** : Development platforms such as Google Colab or Jupyter Notebook for coding and testing.  **Technical Documentation** : Manuals and guides on parameter configuration for different predictive models.  **Training Data** : Data set that will be used to train and validate the models. | 10-17-2024 until 10-18-2024 | Lucas Carrasco | **Difficulties:**  **Parameter Tuning** : Finding optimal values for hyperparameters can be complex and time-consuming. Solution: Perform grid or random searches and use cross-validation techniques to optimize parameters.  **Environment Configuration** : Correctly configuring the development environment and tools can present technical challenges. Solution: Follow specific guides and tutorials to configure your environment and resolve common issues.  **Library Compatibility** : Ensuring compatibility between different versions of libraries can be complicated. Solution: Keep track of versions and use virtual environments to manage dependencies.  **Facilitators:**  **Development Tools** : They facilitate the configuration of parameters and the development of the model efficiently.  **Technical Documentation** : Provides clear guidelines for adjusting parameters and configuring models correctly.  **Prepared Development Environment** : Platforms such as Google Colab allow quick and easy configuration of the development environment, streamlining the process. |
| 1.5.3 | Training of selected models | Training **the Selected Models** involves adjusting the previously chosen predictive models using the prepared data set. This activity includes dividing the data set into training and testing subsets, applying the models to the training data, and adjusting model parameters to optimize model performance. The goal is for models to learn patterns and relationships in the data that allow for accurate predictions. | **Modeling Tools** : Software like Python with libraries like scikit-learn for model training.  **Training Data** : Data set divided into training and testing subsets.  **Technical Documentation** : Guides and manuals on model training and parameter tuning.  **Hardware** : Computers or servers with sufficient processing power to handle model training. | 10-21-2024 until 10-28-2024 | Lucas Carrasco | **Difficulties:**  **Training Time** : The training process can be long, especially with large volumes of data and complex models. Solution: Optimize the code and consider using more powerful computing resources or distributed training techniques.  **Overfitting or Underfitting** : Models can overfit (overfit) or underfit the data. Solution: Tune hyperparameters and use cross-validation techniques to evaluate performance.  **Convergence Problems** : Some models may have difficulty converging to an optimal solution. Solution: Adjust training parameters and use regularization techniques to improve convergence.  **Facilitators:**  **Modeling Tools** : They provide efficient algorithms and methods for training models.  **Technical Documentation** : Helps understand the appropriate methods for training and tuning models.  **Suitable Hardware** : Computers with good processing capacity can speed up the training process and facilitate the management of large volumes of data. |
| 1.5.4 | Model adjustments and optimization | **Model Tuning and Optimization** activity involves refining and improving trained predictive models to maximize their accuracy and effectiveness. This task includes tuning hyperparameters, applying regularization techniques to avoid overfitting, and fine-tuning the models to improve their overall performance. It may also include evaluating different configurations and selecting the models that offer the best balance between accuracy and generalization. | **Optimization Tools** : Software such as Python and specific libraries such as scikit-learn to adjust hyperparameters and apply regularization techniques.  **Validation Data** : Data set used to evaluate the performance of the models during the fitting process.  **Technical Documentation** : Guides on optimization techniques, hyperparameter tuning and model regularization.  **Hardware** : Computers with sufficient processing power to handle the optimization of complex models. | 10-29-2024 until 11-01-2024 | Lucas Carrasco | **Difficulties:**  **Hyperparameter Tuning** : Finding the optimal combination of hyperparameters can be complex and time-consuming. Solution: Use techniques such as grid search or random search to systematize the adjustment.  **Overfitting** : Models can overfit the training data, reducing their ability to generalize. Solution: Apply regularization techniques and use cross-validation to evaluate generalization ability.  **Performance Evaluation** : Determining which metrics to use to evaluate the performance of models can be challenging. Solution: Select appropriate metrics based on project objectives and conduct a thorough evaluation.  **Facilitators:**  **Optimization Tools** : They facilitate the adjustment of hyperparameters and the application of regularization techniques.  **Technical Documentation** : Provides valuable information on best practices for optimizing models and techniques to improve performance.  **Adequate Hardware** : Computers with sufficient capacity can speed up the model tuning and optimization process. |
| 1.5.5 | Performance evaluation using cross validation | **Performance Evaluation Through Cross Validation** consists of measuring the ability of predictive models to generalize to new data and avoid overfitting. This task involves dividing the data set into several subsets (folds) and training the model on some of these subsets while validating on the rest. The results of these evaluations are averaged to obtain a more robust estimate of model performance on unseen data. | **Cross Validation Tools** : Software and libraries such as Python with scikit-learn that provide functions to perform cross validation.  **Training Data** : Data set that will be used to divide into folds and for training and validation of the model.  **Technical Documentation** : Guides on cross-validation techniques and performance metrics.  **Hardware** : Suitable computers to run the cross-validation process, especially if large volumes of data are used. | 04-11-2024 until 07-11-2024 | Lucas Carrasco | **Difficulties:**  **Execution Time** : Cross-validation can be computationally intensive and time-consuming, especially with large data sets and complex models. Solution: Use more efficient cross-validation techniques or reduce data size in initial phases.  **Cross-Validation Overfitting** : Although cross-validation helps evaluate performance, models can still exhibit overfitting. Solution: Complement with other evaluation techniques and adjust hyperparameters as necessary.  **Interpreting Results** : Interpreting performance metrics can be challenging if they are not chosen correctly. Solution: Select appropriate metrics that reflect the project objectives and provide detailed analysis.  **Facilitators:**  **Cross Validation Tools** : They offer automated methods to perform cross validation and evaluate the performance of models efficiently.  **Technical Documentation** : Provides clear guidelines on how to implement and interpret cross-validation.  **Appropriate Hardware** : Computers with sufficient capacity to handle the computational load can facilitate faster and more efficient execution of cross-validation. |
| 1.5.6 | Final adjustments and preparation for evaluation | **Final Adjustments and Preparation for Evaluation** involves making the final modifications to the predictive models and ensuring that they are ready for the final evaluation. This task includes fine-tuning the model parameters based on the results obtained during cross-validation, making minor adjustments to improve the accuracy and generalization of the model, and preparing all documentation and data necessary for the final evaluation of the project. The goal is to ensure that the models are optimized and ready to be presented and evaluated against the project criteria. | **Modeling Tools** : Software and libraries such as Python and scikit-learn to make final adjustments to the models.  **Final Data** : Complete data set that will be used for final adjustments and evaluation.  **Project Documentation** : Information about the modeling process, results obtained and methods used.  **Hardware** : Adequate computers to run and tune the models efficiently. | 07-11-2024 until 08-11-2024 | Lucas Carrasco | **Difficulties:**  **Last-Minute Adjustment** : Making final adjustments can be challenging if unexpected problems arise with models or data. Solution: Perform extensive testing and have a contingency plan for unforeseen problems.  **Complete Documentation** : Ensuring all documentation is complete and clear can be a challenge. Solution: Establish a process for reviewing and verifying documentation before final delivery.  **Preparation for Evaluation** : Preparing all data and results for evaluation can be laborious. Solution: Organize and verify all information in advance to ensure everything is ready for the evaluation.  **Facilitators:**  **Modeling Tools** : They facilitate making final adjustments and preparing models for evaluation.  **Project Documentation** : Helps systematize the process and provide a clear view of the results and methods used.  **Adequate Hardware** : Computers with good processing capacity allow final adjustments and preparations to be made more efficiently. |
| 1.6.1 | Performance measurement | **Performance Measurement** consists of evaluating the effectiveness and precision of the predictive models developed. This activity involves the use of performance metrics, such as precision, recall, F1-score, or R2, to quantify how well the models are meeting established objectives. It also includes the comparison of the results obtained with the benchmarks and success criteria defined in the project to ensure that the models meet expectations. | **Evaluation Tools** : Software and libraries such as Python with scikit-learn to calculate and analyze performance metrics.  **Evaluation Data** : Data set that will be used to measure the performance of the models.  **Technical Documentation** : Guides on performance metrics and model evaluation methods.  **Hardware** : Computers capable of handling the processing necessary for the evaluations. | 11-11-2024 until 11-13-2024 | Lucas Carrasco | **Difficulties:**  **Metric Selection** : Choosing the right metrics can be complicated depending on the project objectives. Solution: Select metrics that best reflect the objectives of the analysis and consult with experts if necessary.  **Interpreting Results** : Interpreting metric results can be challenging if the impact of each metric is not well understood. Solution: Perform a detailed analysis and contextualize the results with the project objective.  **Comparison with Benchmarks** : Comparing performance with benchmarks can be difficult if reference data is not available. Solution: Establish internal benchmarks based on previous models or industry standards.  **Facilitators:**  **Evaluation Tools** : They facilitate the calculation and analysis of performance metrics.  **Technical Documentation** : Provides clear guidelines for measuring and evaluating performance.  **Suitable Hardware** : Computers with good processing capacity allow performance measurements to be carried out efficiently. |
| 1.6.2 | Review of metrics and preliminary results | **Review of Metrics and Preliminary Results** involves analyzing the results obtained during the measurement of the performance of the predictive models. This activity includes verification of calculated metrics, such as precision and recall, and evaluation of initial results to determine whether the models meet established success criteria. The aim is to identify possible inconsistencies, areas for improvement and validate whether the models are aligned with the project objectives. | **Analysis Tools** : Software and libraries such as Python with scikit-learn to review and analyze performance metrics.  **Results Data** : Data generated during model evaluation that will be reviewed to verify results.  **Metric Documentation** : Information on how to interpret and validate the metrics obtained.  **Hardware** : Computers with sufficient capacity to process and analyze performance data. | 11-14-2024 until 11-15-2024 | Lucas Carrasco | **Difficulties:**  **Interpreting Metrics** : It can be difficult to interpret metrics if the results do not meet expectations. Solution: Consult with experts or perform a detailed analysis to better understand the metrics.  **Inconsistency Detection** : Identifying and correcting inconsistencies in results can be challenging. Solution: Conduct thorough reviews and validate data and metrics with multiple sources.  **Model Adjustment** : Adjusting models based on preliminary results may require additional time and adjustments. Solution: Have a clear plan to iterate and optimize models as necessary.  **Facilitators:**  **Analysis Tools** : They facilitate the review and validation of metrics and results.  **Metrics Documentation** : Provides guidelines on how to interpret and validate preliminary results.  **Adequate Hardware** : Computers with good processing capacity allow for efficient review of metrics and results. |
| 1.6.3 | Presentation and discussion of results with stakeholders | **Presentation and Discussion of Results with Stakeholders** involves sharing the findings and results of the predictive analysis with the project stakeholders. This activity includes preparing and delivering a presentation that summarizes key results, performance metrics, and recommendations based on the analysis. The aim is to discuss these results with stakeholders to obtain their feedback, understand their perspectives, and adjust the analysis or models according to their needs and expectations. | **Presentation Tools** : Software such as PowerPoint or Google Slides to create and display the presentation of results.  **Data and Results** : Relevant information and visualizations derived from predictive analysis to be presented.  **Documentation** : Detailed reports and notes that accompany the presentation to explain the results and methodologies used.  **Hardware** : Computers and audiovisual equipment to make the presentation. | 11-18-2024 until 11-20-2024 | Lucas Carrasco | **Difficulties:**  **Communicating Results** : Explaining technical results to a non-technical audience can be challenging. Solution: Simplify findings and use clear visualizations to facilitate understanding.  **Receiving Feedback** : Getting constructive feedback and handling potential criticism can be complicated. Solution: Prepare for questions and be open to suggestions, using feedback to adjust the analysis.  **Later Adjustments** : Implementing changes based on the discussion may require additional time and adjustments. Solution: Establish a clear plan for making modifications and updating models as necessary.  **Facilitators:**  **Presentation Tools** : Make it easy to create and deliver a professional and clear presentation.  **Documentation** : Provides additional support to explain and contextualize the results presented.  **Adequate Hardware** : Allows the presentation to be carried out effectively, with adequate audiovisual support. |
| 1.6.4 | Adjustments to results based on feedback | **Adjusting Results Based on Feedback** involves modifying and optimizing results and predictive models in response to comments and suggestions received during presentation to stakeholders. This activity includes reviewing the feedback provided, identifying areas for improvement, and implementing necessary adjustments to the models and analyzes to better align them with the expectations and needs of the project. The goal is to ensure that the results are more accurate, relevant and useful to stakeholders. | **Analysis Tools** : Software and libraries such as Python and scikit-learn to adjust and optimize predictive models.  **Stakeholder Feedback** : Comments and suggestions received during the presentation that will guide adjustments.  **Model Documentation** : Information about current models and results to facilitate the implementation of changes.  **Hardware** : Appropriate computers to carry out the necessary modifications and tests. | 11-21-2024 until 11-22-2024 | Lucas Carrasco | **Difficulties:**  **Implementing Changes** : Making adjustments to models can be complex and time-consuming. Solution: Prioritize the most critical changes and plan a structured approach to their implementation.  **Feedback Integration** : Aligning changes with suggestions can be challenging if the feedback is broad or contradictory. Solution: Evaluate the feedback carefully and decide the adjustments that will offer the most value.  **Adjustment Validation** : Verifying that adjustments actually improve results may require additional time and retesting. Solution: Perform extensive testing to validate changes and ensure improved results are consistent.  **Facilitators:**  **Stakeholder Feedback** : Provides valuable perspective to guide adjustments and improvements.  **Analysis Tools** : They facilitate the implementation and testing of changes made to the models.  **Model Documentation** : Helps understand the necessary changes and implement adjustments efficiently. |
| 1.7.1 | Deployment of the model in google colab | **Model Deployment on Google Colab** involves deploying and running the final predictive model on Google Colab, a cloud notebook platform. This activity includes preparing the Google Colab environment, loading data and model code, and running the model to validate its performance in a real environment. The goal is to ensure that the model runs correctly in Google Colab and that the results are consistent with those obtained in the initial development. | **Google Colab** : Cloud notebook platform to run model code.  **Model Code** : Scripts and code cells that define and train the predictive model.  **Data** : Data sets required to test and validate the model in the Google Colab environment.  **Documentation** : Instructions and comments in the code to facilitate deployment and execution in Google Colab.  **Hardware** : Computer with internet access to work on Google Colab. | 11-25-2024 until 11-26-2024 | Lucas Carrasco | **Difficulties:**  **Environment Setup** : It can be difficult to set up the environment in Google Colab if there are issues with code compatibility or data loading. Solution: Review and adjust the code according to the needs of the Colab environment and follow the configuration guides.  **Model Execution** : Large or complex models may experience performance or runtime issues in Google Colab. Solution: Optimize the code and split tasks if necessary to improve efficiency.  **Resource Management** : Google Colab has resource limitations that can affect model performance. Solution: Monitor resource usage and adjust data size or model complexity to accommodate constraints.  **Facilitators:**  **Access to Google Colab** : Allows you to run code in the cloud without the need for local infrastructure, facilitating access to tools and resources.  **Documentation and Tutorials** : Google Colab offers extensive documentation and tutorials that can help in configuring and deploying the model.  **Ease of Sharing** : Google Colab allows you to easily share notebooks with other team members for review and collaboration. |
| 1.7.2 | Testing and tuning in the production environment | **Testing and Tuning in the Production Environment** involves validating and tuning the predictive model in a real production environment. This activity includes deploying the model to the production environment, running tests to ensure its correct operation, and making adjustments based on the results of these tests. The goal is to ensure that the model works effectively and efficiently in the environment in which it will be used, and to make the necessary adjustments to optimize its performance. | **Production Environment** : Infrastructure and systems where the model will be deployed for real use.  **Model Code** : Final versions of the predictive model and scripts necessary for its execution in production.  **Test Data** : Representative data that will be used to test the model in the production environment.  **Monitoring Tools** : Software to monitor model performance and detect problems.  **Documentation** : Guides and records on the model, test procedures and adjustment criteria. | 11-27-2024 until 11-27-2024 | Lucas Carrasco | **Difficulties:**  **Integration with Existing Systems** : There may be problems integrating the model with other systems and processes in the production environment. Solution: Perform extensive integration testing and collaborate with the IT team to resolve issues.  **Model Performance** : The model may not perform as expected due to variations in the data or environment. Solution: Implement adjustments and optimizations based on test results and continuous monitoring.  **Resource Management** : The resources available in the production environment may be limited, affecting the performance of the model. Solution: Optimize the code and tune the model to fit resource constraints.  **Facilitators:**  **Controlled Environment** : Allows you to test the model in real conditions and obtain accurate feedback.  **Monitoring Tools** : They facilitate model monitoring and problem detection in real time.  **Documentation and Procedures** : They offer clear guidance for testing and adjusting the model, ensuring effective and organized execution. |
| 1.7.3 | Creation of technical documentation | **Creation of Technical Documentation** involves developing and compiling all the necessary documentation that explains the design, development, implementation and operation of the predictive model. This activity includes the preparation of user manuals, implementation guides, and technical documents that describe the algorithms used, the processes followed, and the data structures. The goal is to provide a complete and clear reference for future users, maintainers and developers of the system, ensuring that the model is understandable and usable in the long term. | **Documentation Tools** : Software such as Microsoft Word, Google Docs, or specialized tools for technical documentation.  **Technical Information** : Details about the model design, algorithms used, and processed data.  **User Manual** : Guides and manuals that explain how to use the model and its results.  **Development Documentation** : Records of design decisions, source code, and implementation procedures. | 11-28-2024 until 11-28-2024 | Lucas Carrasco | **Difficulties:**  **Complexity of Information** : Documentation can become extensive and technical, making it difficult to write. Solution: Divide the documentation into clear sections and use accessible language.  **Continuous Update** : Keeping documentation up to date with model changes can be a challenge. Solution: Implement a process for regular review and updating of documentation.  **Determining Scope** : Defining what information to include can be complicated. Solution: Follow a standard outline and ensure that all essential aspects of the model are covered.  **Facilitators:**  **Templates and Guides** : Use of technical documentation templates and guides that facilitate the structure and format of the content.  **Collaboration** : Teamwork to review and complete documentation, ensuring all relevant information is included.  **Previous Experience** : Experience creating technical documentation can speed up the process and improve the quality of the documents. |
| 1.7.4 | Preparation of the user guide | **Preparing the User Guide** consists of developing a clear and accessible document that instructs users on how to interact with the predictive model. This activity includes writing detailed instructions on how to use the system, interpret the results, and perform common tasks. The goal is to ensure that users can use the model effectively, without requiring constant technical support, and that they understand how to get the most out of the available functionality. |  | 11-29-2024 until 11-29-2024 | Lucas Carrasco | **Difficulties:**  **System Complexity** : Explaining complex functionalities can be challenging. Solution: Divide the guide into sections and use simple language and clear examples.  **User Diversity** : Different levels of user experience can complicate writing. Solution: Include basic and advanced sections to cover the needs of all users.  **Updating the Guide** : Keeping the guide up to date with changes to the model can be complicated. Solution: Implement a regular review and update process.  **Facilitators:**  **Guide Templates** : Use of pre-existing templates that facilitate the structuring of the content.  **User Feedback** : Receiving feedback from users during testing can help improve the clarity and usefulness of the guide.  **Previous Experience** : Experience in writing manuals and guides can speed up the creation of an effective and useful document. |
| 1.8.1 | Monitoring the performance of the model in production | **Monitoring Model Performance in Production** involves monitoring the performance of the predictive model after its deployment in the production environment. This activity includes collecting data on the performance of the model, evaluating its accuracy and efficiency, and identifying any problems or deviations in performance. The goal is to ensure that the model continues to operate as expected, optimize its performance, and make adjustments as necessary to maintain the quality and reliability of the predictive analysis. | **Monitoring Tools** : Software for monitoring model performance and data management.  **Production Data** : Real information from the production environment to evaluate the model.  **Evaluation Protocols** : Methods and metrics to measure the precision, efficiency and effectiveness of the model.  **Support Teams** : Technical staff to resolve problems and make adjustments to the model. | 02-12-2024 until 03-12-2024 | Lucas Carrasco | **Difficulties:**  **Performance Deviations** : The model may present problems or inconsistencies in the production environment. Solution: Implement an alert system and make adjustments and recalibrations as necessary.  **Data Volume** : Managing large volumes of data can be a challenge. Solution: Use efficient data processing and optimization tools.  **Technical Resources** : Limitations in available resources may affect monitoring capacity. Solution: Ensure the production environment is adequately equipped to handle the demands of the model.  **Facilitators:**  **Advanced Monitoring Tools** : Facilitate monitoring and analysis of model performance in real time.  **Clear Metrics** : Defining accurate and relevant performance metrics allows for effective evaluation.  **Technical Support** : A trained technical team can quickly resolve problems and keep the model in optimal condition. |
| 1.8.2 | Initial feedback collection and adjustment | **Feedback Collection and Initial Adjustments** involves collecting opinions and comments from users and stakeholders about the performance of the predictive model after its implementation. This activity includes evaluating user satisfaction, identifying areas for improvement, and making initial adjustments to optimize the model. The goal is to ensure that the model adapts to the real needs of users and that it works effectively in the production environment. | **Survey Tools** : Platforms to collect feedback, such as Google Forms or online surveys.  **Communication Channels** : Means of collecting feedback, such as emails, meetings and forums.  **Feedback Data** : Opinions and suggestions from users and stakeholders.  **Development Team** : Personnel in charge of reviewing the feedback and making the necessary adjustments. | 04-12-2024 until 04-12-2024 | Lucas Carrasco | **Difficulties:**  **Feedback Quality** : Feedback may be inaccurate or difficult to interpret. Solution: Use specific, structured questions in surveys to get clear feedback.  **Time to Implement Adjustments** : Making adjustments can take time and resources. Solution: Prioritize adjustments based on the impact and urgency of the feedback received.  **Effective Communication** : Coordinating with users and stakeholders to receive feedback can be a challenge. Solution: Establish clear and accessible communication channels.  **Facilitators:**  **Effective Surveys** : Well-designed survey tools make it easy to collect valuable feedback.  **Proactive Collaboration** : The cooperation of users and stakeholders can improve the quality of feedback and the effectiveness of adjustments.  **Support Team** : A dedicated review and adjustment team can implement changes efficiently and quickly. |
| 1.8.3 | Making necessary adjustments and updating the model | **Making Necessary Adjustments and Model Updates** consists of implementing modifications and improvements to the predictive model based on the feedback collected and continuous analysis of performance. This activity includes adjusting model parameters, updating algorithms or techniques used, and making changes to the data structure as necessary to optimize model performance. The goal is to ensure that the model remains relevant and effective over time, adapting to new conditions or requirements. | **Modeling Tools** : Software and platforms such as Google Colab or Jupyter Notebook to adjust and update the model.  **Updated Data** : Recent information to retrain the model and reflect changes in the environment.  **Technical Documentation** : Manuals and records on the adjustments made and the versions of the model.  **Development Team** : Personnel with programming and data analysis skills to implement the necessary adjustments. | 05-12-2024 until 05-12-2024 | Lucas Carrasco | **Difficulties:**  **Data Compatibility** : Ensuring that new data is compatible with the model can be complicated. Solution: Perform extensive testing and validation before upgrading.  **Adjustment Time** : Adjustments to the model can take time and temporarily affect performance. Solution: Plan maintenance and testing periods to minimize impact.  **Errors in Implementation** : Adjustments could introduce new errors or problems. Solution: Implement changes in stages and perform rigorous testing to identify and correct problems.  **Facilitators:**  **Detailed Documentation** : Having clear documentation about the model and settings makes implementation and troubleshooting easier.  **Development Tools** : The use of advanced tools for adjusting and updating the model can speed up the process.  **Continuous Feedback** : Continuous feedback from users and stakeholders helps to make more precise and effective adjustments. |
| 1.8.4 | Final report preparation and closing | **Preparation of the Final Report and Closure** involves compiling all relevant information and results obtained during the project into a detailed report. This activity includes writing the final report, presenting the findings, documenting the processes carried out and evaluating the project results. The goal is to provide a clear and complete summary of the project, evaluate its success, and formalize project closure. | **Documentation Tools** : Word processing software such as Microsoft Word or Google Docs to write the report.  **Data and Results** : Information about the results of the analysis, adjustments made and performance metrics.  **Report Templates** : Structures and formats for the presentation of the final report.  **Review Team** : People in charge of reviewing and approving the final report, such as supervisors or colleagues. | 06-12-2024 until 06-12-2024 | Lucas Carrasco | **Difficulties:**  **Information Gathering** : Gathering and organizing all the relevant information can be a challenge. Solution: Maintain detailed and up-to-date records throughout the project.  **Clarity in Presentation** : Ensuring that the report is clear and understandable can be complicated. Solution: Use a structured format and receive feedback from reviewers.  **Meeting Deadlines** : Completing the final report within the established deadlines can be stressful. Solution: Plan ahead and allocate enough time for writing and revising.  **Facilitators:**  **Continuous Documentation** : Maintaining detailed documentation throughout the project facilitates the preparation of the final report.  **Templates and Guides** : Using templates and guides for writing the report can speed up the process.  **Support Team** : Have a review and support team to provide feedback and ensure the quality of the report. |

| 8. Gantt chart |
| --- |
| Find a Gantt Chart format that suits you and organize in it the activities planned in the previous point considering the period assigned for the development of your APT Project. You must maintain the temporality of the academic period in the development of the three phases contemplated by the Title Portfolio Subject. |

**Total Project Duration: 15 weeks:** This schedule considers a logical sequence of activities with reasonable times for each phase, ensuring that all key stages of the project are completed within the established deadline, with room for adjustments and validations.

****



1. If the APT Project is a group project, in this column you must indicate the name of those responsible for each task or activity. This will later allow the evaluation to be differentiated for each member. [↑](#footnote-ref-0)