Salifort Motors PACE Strategy Document

Workforce Retention Analysis

Introduction

PACE stands for Plan, Analyze, Construct and Execute. It is a framework that illustrates the foundation and structure for data analysis projects and each letter represents an actionable stage in a project. The stage “Plan” involves the definition of the project scope, the research of business data and the workflow development. The stage “Analyze” involves data scrubbing, data conversion and database formatting. The stage “Construct” involves building models and machine learning algorithms and selecting a modeling approach. The stage “Execute” involves the presentation of results to decision-makers, stakeholders and others in order to receive feedback. This framework is built upon an iterative cycle where each stage may reveal new insights, requiring the return to earlier stages. A PACE strategy document is used to record decisions and reflections at different stages of the data analytical process. It typically includes the definitions of roles and actions to ensure clarity and accountability.

Purpose

Salifort Motors' HR department aims to enhance employee satisfaction. The data has been collected but guidance is needed on how to use it effectively. We want to extract insights from the data and provide evidence-based recommendations to the department. Their key concern is understanding what factors might cause an employee to leave the company. The goal of this project is to analyze the HR data and build a predictive model that determines whether an employee will leave the company. By identifying those at risk of leaving, the company can better address retention, which can save the time and cost of recruitment and training.

Project proposal

Salifort Motors Project Proposal

Overview

Salifort Motors's concern is understanding what factors might cause an employee to leave the company. The goal of this project is to analyze the HR data and establish a method that determines whether an employee will leave the company.

Milestones	Tasks	PACE stages
1	Comprehension of Business Context & Problem Definition	Plan
2	Data Investigation & Cleaning	Plan, Analyze
3	Model Selection	Analyze, Construct
4	Model Construction	Construct
5	Validation of Model Assumptions	Analyze, Construct
6	Model Evaluation	Analyze
7	Result Interpretation & Presentation to Stakeholders	Execute

Considerations

PACE: Plan Stage

Data Science

- Who is the intended audience for this project?
- What problem are we trying to address and how do we expect this work to impact the broader business objective?
- Which questions need to be explored or resolved?
- What resources are necessary to complete the project?
- What are the key deliverables throughout this project?

Python

- How can we best organize and comprehend the information provided?
- Which codebooks for practice and review will assist in this task?
- What preliminary steps should a proactive learner take before beginning to code?

Data Translation into Insights

- What are the data columns and variables and which ones are most critical for the deliverable?
- What units are used for each variable?
- What assumptions do we have about the data that could guide our exploratory analysis, keeping in mind these will need to be validated?
- Is there any data that is missing or incomplete?
- Are all parts of the dataset in a consistent format?
- Which Exploratory Data Analysis (EDA) methods will be essential to initiate this project?

Statistics

- What is the primary objective of this project?
- What is the research question for this study?
- Why is random sampling important and what could be an example of sampling bias if random sampling is not used?

Regression Analysis

- Who are the stakeholders involved in this project?
- What problem are we aiming to solve?

- What are our initial insights from examining the data?
- Which resources are we utilizing during this phase?
- Are there any ethical considerations at this stage?

Machine Learning

- What problem are we addressing?
- Which resources are helpful as we progress through this stage?
- Is the data reliable?
- Are there any new ethical concerns at this point?
- What ideal data would we need to fully address this question?
- What data is currently available or accessible?
- What metric should be used to measure the success of the business goal and why?

Data Project Questions & Considerations

PACE: Analyze Stage

Python

- Will the available data be adequate to achieve the goal, based on our understanding and variable analysis?

Data Translation into Insights

- What are the necessary steps for conducting EDA to effectively meet the project objectives?
- Is there a need to supplement the data by joining datasets? What structuring is required (e.g. filtering, sorting?
- What initial assumptions do we have regarding suitable visualizations for our audience?

Statistics

- How can descriptive statistics be beneficial?
- How do the null hypothesis and the alternative hypothesis differ?

Regression Analysis

- Why is EDA valuable before building a linear regression model?
- Are there any ethical concerns to consider during this phase?

Machine Learning

- What problem are we addressing and does the plan still align with the goal or does it need adjustments?
- Does the data meet the model's assumptions and, if not, is it acceptable or problematic?
- Why did we choose the specific independent variables?
- What is the role of EDA before developing a model?
- What insights have we gained from the EDA?
- What resources are we using in this phase?
- Are there any ethical concerns during this stage?

Data Project Questions & Considerations

PACE: Construct Stage

Python

- Are any of the data variable averages unusual?
- How many vendors, organizations or groups are included in the dataset?

Data Translation into Insights

- What visualizations, machine learning models or other outputs need to be created to achieve the project goals?
- What processes are required to build the necessary visualizations?
- Which variables are most relevant for visualizations in this project?
- Reflecting on the Plan stage, how do we plan to handle missing data?

Statistics

- How did we develop our null and alternative hypotheses?
- What conclusions can be drawn from the hypothesis test?

Regression Analysis

- Did we notice any irregularities?
- Can improvements be made? What would we change about the model?

Machine Learning

- Is there a problem? If so, how can it be resolved?
- Which independent variables were selected for the model and why?
- How well does the model align with the data (e.g. validation score)?
- Can improvements be made? What adjustments would we make to the model?
- Are there any ethical concerns in this phase?

Data Project Questions & Considerations

PACE: Execute Stage

Python

- Based on our current understanding of the data, what would we suggest investigating further before conducting an exploratory data analysis (EDA)?
- Which part of the data initially appears to contain anomalies?
- What additional data could strengthen this dataset?

Data Translation into Insights

- What key findings have come from our EDA and visualizations?
- What business recommendations can we make based on the visualizations?
- Given our understanding of the data and visualizations, what other questions could be explored?
- How would we present these visualizations to different audiences?

Statistics

- What important business insights emerged from our A/B testing?
- What recommendations can we make based on our findings?

Regression Analysis

- Why is it crucial to interpret beta coefficients when analyzing model results?
- What potential recommendations could we offer to our manager or company?
- Could the model be enhanced? Why or why not?
- What business suggestions can we make based on the models developed?
- What key insights have emerged from our models?
- Are there any ethical considerations in this phase?

Machine Learning

- What key insights have emerged from the models?
- What factors should guide model selection?

- Does the model make sense and are the final outcomes acceptable?
- Were there any features that turned out to be insignificant? What happens if they are removed?
- Based on our understanding of the data and models, what other questions could be explored?
- What resources are we using during this stage?
- Is the model ethically sound?
- When the model makes errors, what might be the cause and how does this relate to the specific use case?