.

# Portfolio Project Recap

* Understand the data in the problem context.
* Consider how the data will best address the business need.
* Contextualize and understand the data and the problem.
* Perform EDA (understand the variables and analyze relationships between them)
* Create visualizations.
* Determine which models are most appropriate.
* Construct the model.
* Confirm model assumptions.
* Evaluate model results to determine how well the model fits the data.
* Interpret model performance and results.
* Share actionable steps with stakeholders.

**Project proposal**

**Salifort Motors project proposal**

## **Overview**

*Salifort Motors is seeking a method to use employee data to gauge what makes them leave the company.*

|  |  |  |
| --- | --- | --- |
| **Milestones** | **Tasks** | **PACE stages** |
| **1** | **Understand the business scenario and define the problem** | **Plan** |
| **2** | **Data exploration and data cleaning** | **Plan, Analyze** |
| **3** | **Determine which models are most appropriate** | **Analyze,Construct** |
| **4** | **Construct the model** | **Construct** |
| **5** | **Confirm model assumptions** | **Analyze, Construct** |
| **6** | **Evaluate model results** | **Analyze** |
| **7** | **Interpret results and share actionable steps with stakeholders** | **Execute** |

**Data Project Questions & Considerations**

**PACE: Plan Stage**

**Python**

* How can I best prepare to understand and organize the provided information?
* What follow-along and self-review codebooks will help you perform this work?

**Translate Data into Insights**

* What are the data columns and variables, and which ones are most relevant to my deliverable?
* What units are my variables in?
* What are my initial presumptions about the data that can inform your EDA, knowing I will need to confirm or deny your future findings?
* Is there any missing or incomplete data?
* Are all pieces of this dataset in the same format?
* Which EDA practices will be required to begin this project?

**Statistical Analysis**

* What is the main purpose of this project?
* What is the research question for this project?
* What is the importance of random sampling? In this case, what is an example of sampling bias that might occur if you didn’t use random sampling?

**Regression Analysis**

* Who are my stakeholders for this project?
* What am I trying to solve or accomplish?
* What are my initial observations when I explore the data?
* What resources do you find myself using as I complete this stage?
* Do I have any ethical considerations at this stage?

**Machine Learning Models**

* What am I trying to solve?
* What resources do I find myself using as I complete this stage?
* Is my data reliable?
* Do I have any additional ethical considerations in this stage?
* What data do I need/would I like to see in a perfect world to answer this question?
* What data do I have/can I get?
* What metric should I use to evaluate the success of my business objective? Why?

**Data Project Questions & Considerations**

**PACE: Analyze Stage**

**Python**

* Will the available information be sufficient to achieve the goal based on my intuition and the analysis of the variables?

**Translate Data into Insights**

* What steps need to be taken to perform EDA in the most effective way to achieve the project goal?
* Do I need to add more data using the EDA practice of joining? What type of structuring needs to be done to this dataset, such as filtering, sorting, etc.?
* What initial assumptions do I have about the types of visualizations that might best be suited for the intended audience?

**Statistical Analysis**

* Why are descriptive statistics useful?
* What is the difference between the null hypothesis and the alternative hypothesis?

**Regression Analysis**

* What are some purposes of EDA before constructing a multiple linear regression model?
* Do I have any ethical considerations at this stage?

**Machine Learning Models**

* What am I trying to solve? Does it still work? Does the plan need revising?
* Does the data break the assumptions of the model? Is that ok, or unacceptable?
* Why did I select the X variables?
* What are some purposes of EDA before constructing a model?
* What has the EDA told me?
* What resources do I find myself using as I complete this stage?
* Do I have any ethical considerations in this stage?

**PACE: Construct Stage**

**Python**

* Do any data variable averages look unusual?
* How many vendors, organizations or groupings are included in this total data?

**Translate Data into Insights**

* What data visualizations, machine learning algorithms, or other data outputs will need to be built in order to complete the project goals?
* What processes need to be performed in order to build the necessary data visualizations?
* Which variables are most applicable for the visualizations in this data project?
* Going back to the Plan stage, how do I plan to deal with the missing data (if any)?

**Statistical Analysis**

* How did I formulate my null hypothesis and alternative hypothesis?
* What conclusion can be drawn from the hypothesis test?

**Regression Analysis**

* Do I notice anything odd?
* Can I improve it? Is there anything I would change about the model?

**Machine Learning Models**

* Is there a problem? Can it be fixed? If so, how?
* Which independent variables did I choose for the model, and why?
* How well does my model fit the data? (What is my model’s validation score?)
* Can I improve it? Is there anything I would change about the model?
* Do I have any ethical considerations at this stage?

**PACE: Execute Stage**

**Python**

* Given my current knowledge of the data, what would I initially recommend to my manager to investigate further before performing an exploratory data analysis?
* What data initially presents as containing anomalies?
* What additional types of data could strengthen this dataset?

**Translate Data into Insights**

* What key insights emerged from my EDA and visualizations(s)?
* What business recommendations do I propose based on the visualization(s) built?
* Given what I know about the data and the visualizations I was using, what other questions could I research for the team?
* How might I share these visualizations with different audiences?

**Statistical Analysis**

* What key business insight(s) emerged from my A/B test?
* What business recommendations do I propose based on my results?

**Regression Analysis**

* To interpret model results, why is it important to interpret the beta coefficients?
* What potential recommendations would I make to my manager/company?
* Do I think my model could be improved? Why or why not? How?
* What business recommendations do I propose based on the models built?
* What key insights emerged from my model(s)?
* Do I have any ethical considerations at this stage?

**Machine Learning Models**

* What key insights emerged from my model(s)?
* What are the criteria for model selection?
* Does my model make sense? Are my final results acceptable?
* Were there any features that were not important at all? What if I take them out?
* Given what I know about the data and the models I was using, what other questions could I address for the team?
* What resources do I find myself using as I complete this stage?
* Is my model ethical?
* When my model makes a mistake, what is happening? How does that translate to my use case?