**Capstone project scenario overview: Working for Salifort Motors**

**About the company**

Salifort Motors is a fictional French-based alternative energy vehicle manufacturer. Its global workforce of over 100,000 employees research, design, construct, validate, and distribute electric, solar, algae, and hydrogen-based vehicles. Salifort’s end-to-end vertical integration model has made it a global leader at the intersection of alternative energy and automobiles.

**Your business case**

As a data specialist working for Salifort Motors, you have received the results of a recent employee survey. The senior leadership team has tasked you with analyzing the data to come up with ideas for how to increase employee retention. To help with this, they would like you to design a model that predicts whether an employee will leave the company based on their  department, number of projects, average monthly hours, and any other data points you deem helpful.

**The value of your deliverable**

For this deliverable, you are asked to choose a method to approach this data challenge based on your prior course work. Select either a regression model or a machine learning model to predict whether an employee will leave the company. Both approaches are shown in the project exemplar, but only one is needed to complete your project.

The capstone project will provide you with valuable experience and data analysis examples that you can share with potential employers.

**Key takeaways**

Your capstone project will enable you to apply your newly developed advanced data analytics skills and knowledge, demonstrate fundamental data analysis skills to prospective employers, and present what you have learned from the Grow with Google Advanced Data Analytics Certificate. Providing examples from projects such as the capstone project is an excellent way to showcase your data analytics abilities. The end results from the capstone project greatly strengthen your resume and make you a more competitive candidate for data analytics positions.

**Activity Overview**

In this activity, you will showcase your ability to use Python for model building and data analysis. You will deploy different models to analyze a dataset and generate business insights for your stakeholders. In particular, you will build and evaluate a logistic regression model or the following machine learning models: decision tree, random forest, XGBoost. You will also update your stakeholders through an executive summary, demonstrating your ability to organize and communicate key information.

**Scenario**

You are a data professional working for Salifort Motors.

Currently, there is a high rate of turnover among Salifort employees. (Note: In this context, turnover data includes both employees who choose to quit their job and employees who are let go). Salifort’s senior leadership team is concerned about how many employees are leaving the company. Salifort strives to create a corporate culture that supports employee success and professional development. Further, the high turnover rate is costly in the financial sense. Salifort makes a big investment in recruiting, training, and upskilling its employees.

If Salifort could predict whether an employee will leave the company, and discover the reasons behind their departure, they could better understand the problem and develop a solution.

As a first step, the leadership team asks Human Resources to survey a sample of employees to learn more about what might be driving turnover.

Next, the leadership team asks you to analyze the survey data and come up with ideas for how to increase employee retention. To help with this, they suggest you design a model that predicts whether an employee will leave the company based on their job title, department, number of projects, average monthly hours, and any other relevant data points. A good model will help the company increase retention and job satisfaction for current employees, and save money and time training new employees.

As a specialist in data analysis, the leadership team leaves it up to you to choose an approach for building the most effective model to predict employee departure. For example, you could build and evaluate a statistical model such as logistic regression. Or, you could build and evaluate machine learning models such as decision tree, random forest, and XGBoost. Or, you could choose to deploy both statistical and machine learning models.

For any approach, you’ll need to analyze the key factors driving employee turnover, build an effective model, and share recommendations for next steps with the leadership team.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Step-By-Step Instructions**

Follow the instructions to complete the activity. Then, go to the next course item to compare your work to a completed exemplar.

Step 1: Access the templates

To use the templates for this course item, click each link below and select *Use Template*.

[Activity Template\_ Course 7 PACE strategy document](https://d3c33hcgiwev3.cloudfront.net/g6FgRcIESsqaIfLuPFJzSg_431833d72f434614af88cf1d8f4d12f1_Activity-Template_-Course-7-PACE-strategy-document.docx?Expires=1687996800&Signature=GWfeJQEMF~jfUkPD~DDbu4cUt2l4fCrc4AzJEu1M8MxY1jxOi3m1Rcok9kBvBqZNjFs~vOrn-jkWFB7sLd4HLwj3KOO485Vk0UZLfRraSSqGDET9Cu3L5TrvpiriNCe26KpqgJzy2VQ83XMcCJpmxKUCcjiTidcBmsbocya3fUk_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A" \t "_blank)

[DOCX File](https://d3c33hcgiwev3.cloudfront.net/g6FgRcIESsqaIfLuPFJzSg_431833d72f434614af88cf1d8f4d12f1_Activity-Template_-Course-7-PACE-strategy-document.docx?Expires=1687996800&Signature=GWfeJQEMF~jfUkPD~DDbu4cUt2l4fCrc4AzJEu1M8MxY1jxOi3m1Rcok9kBvBqZNjFs~vOrn-jkWFB7sLd4HLwj3KOO485Vk0UZLfRraSSqGDET9Cu3L5TrvpiriNCe26KpqgJzy2VQ83XMcCJpmxKUCcjiTidcBmsbocya3fUk_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A" \t "_blank)

[Activity Templates\_ Executive summaries](https://d3c33hcgiwev3.cloudfront.net/dsJAWJx-Q-COlR03Tm64ew_c820170fbfc840bca08d6b234846caf1_Activity-Templates_-Executive-summaries.pptx?Expires=1687996800&Signature=eLYnJDNxY-CX~~NDiS03mbXBGXsSB2l2TyaJaHJa7od2m8sUGd9fJu70L08qPC-mFTp0j3AlkbyQTNp451e78tn21G7q9kjqv~ngzto0BMAlNnWTgYdtdcxGxCtKyKQU4hG5hPFBxFuXXDj8XilneoaSNS9gIWpniB4xwi6ZlnU_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A" \t "_blank)

[PPTX File](https://d3c33hcgiwev3.cloudfront.net/dsJAWJx-Q-COlR03Tm64ew_c820170fbfc840bca08d6b234846caf1_Activity-Templates_-Executive-summaries.pptx?Expires=1687996800&Signature=eLYnJDNxY-CX~~NDiS03mbXBGXsSB2l2TyaJaHJa7od2m8sUGd9fJu70L08qPC-mFTp0j3AlkbyQTNp451e78tn21G7q9kjqv~ngzto0BMAlNnWTgYdtdcxGxCtKyKQU4hG5hPFBxFuXXDj8XilneoaSNS9gIWpniB4xwi6ZlnU_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A" \t "_blank)

Step 2: Access the capstone project lab

***Note****: The following lab is also the next course item. Once you complete and submit your activity, return to the lab instructions’ page and click* ***Next*** *to continue on to the exemplar reading.*

To access the lab, click the link below and select *Open Lab:*

* [Course 7 capstone lab](https://www.coursera.org/learn/google-advanced-data-analytics-capstone/ungradedLab/uskOX/activity-course-7-salifort-motors-project-lab)

Your Python notebook for this project includes a guided framework that will assist you with the required coding. Input the code and answer the questions in your Python notebook to build and evaluate a logistic regression model or the following machine learning models: decision tree, random forest, XGBoost. You’ll find helpful reminders for tasks like:

* Exploratory data analysis (EDA)
* Model building and evaluation

You will also discover questions in this Python notebook designed to help you gather the relevant information you’ll need to write an executive summary for your stakeholders.

Use your completed PACE strategy document and Python notebook to help you prepare your executive summary.

Data Dictionary

This project uses a dataset called **HR\_capstone\_dataset.csv.** It represents 10 columns of self-reported information from employees of a multinational vehicle manufacturing corporation.

The dataset contains:

14,999 rows – each row is a different employee’s self-reported information

10 columns

| **Column name** | **Type** | **Description** |
| --- | --- | --- |
| satisfaction\_level | int64 | The employee’s self-reported satisfaction level [0-1] |
| last\_evaluation | int64 | Score of employee's last performance review [0–1] |
| number\_project | int64 | Number of projects employee contributes to |
| average\_monthly\_hours | int64 | Average number of hours employee worked per month |
| time\_spend\_company | int64 | How long the employee has been with the company (years) |
| work\_accident | int64 | Whether or not the employee experienced an accident while at work |
| left | int64 | Whether or not the employee left the company |
| promotion\_last\_5years | int64 | Whether or not the employee was promoted in the last 5 years |
| department | str | The employee's department |
| salary | str | The employee's salary (low, medium, or high) |
| satisfaction\_level | int64 | The employee’s self-reported satisfaction level [0-1] |
| last\_evaluation | int64 | Score of employee's last performance review [0–1] |

Step 3: Complete your PACE strategy document

The **Capstone PACE strategy document** includes questions that will help guide you through the Course 7 employee data project. Answer the questions in your PACE strategy document to prepare for using Python to build and evaluate statistical, regression, and/or machine learning model(s) to analyze your data.

As a reminder, the PACE strategy document is designed to help you complete the contents for each of the templates provided. You may navigate back and forth between the PACE strategy document and the Python notebook as needed.

Step 4: Prepare an executive summary

Your executive summary will keep the senior leadership team at Salifort Motors informed of your progress. The one-page format is designed to respect teammates and stakeholders who might not have time to read and understand a lengthy technical report.

First, select one of the executive summary design layouts from the provided template.

Then, add the relevant information. Your executive summary should include the following:

* A summary of the benefits and limitations of your regression, or machine learning model(s)
* The results of your analysis
* Recommendations or insights based on your results, including recommended next steps

Finally, complete your executive summary to effectively communicate your findings to your stakeholders. Ensure that the model’s limitations are addressed, model results are clearly stated, and next steps are identified.

**Pro Tip: Save the templates**

Finally, be sure to save a blank copy of the templates you used to complete this activity. You can use them for further practice or in your professional projects. These templates will help you work through your thought processes and demonstrate your experience to potential employers.

**What to Include in Your Response**



**Course 7 PACE strategy document:**

* Answer the questions in the PACE strategy document

**Course 7 capstone lab:**

* Build and evaluate a logistic regression model

or

* Build and evaluate the following machine learning models: decision tree, random forest, XGBoost

**Course 7 executive summary:**

* Summarize the benefits and limitations of your regression, or machine learning model(s)
* Identify the results of your analysis
* Include recommendations or insights based on your results

**Course 7 PACE strategy document**

Compare the exemplar to your completed PACE strategy document. Your responses might differ in some respects from the exemplar, but that is to be expected. What did you do well? Where can you improve? Use your answers to these questions to guide you as you complete the capstone project.

***Note:*** *The exemplar represents one possible way to address the PACE questions and classify project tasks according to the PACE workflow. Yours might differ in certain ways, such as your specific responses to questions. What’s important is that you have a general idea of the order and classification of tasks in the data analysis process.*

Your PACE strategy document should:

* Outline an effective plan of action for your project based on the PACE structure
* Clearly communicate the reasoning behind your decisions to pair PACE stages with specific tasks



**Course 7 capstone lab**

Compare your notebook against the exemplar notebook provided. Your responses might differ in some respects from the exemplar, but that is to be expected. What did you do well? Where can you improve? Use your answers to these questions to guide you as you continue to progress through the course.

***Note:*** *Earlier in this certificate, you completed one or several portfolio projects. Each portfolio project included a project exemplar, which represented one possible way to complete the Python notebook. Your project might have differed in certain ways, such as your specific code or responses to questions. What's important is that you have gained an overall understanding of the purpose and functionality of a Python notebook for data analysis.* ***However,*** *for this Course 7 Capstone project, the Python lab notebook showcases many approaches to find a solution to the presented business problem. In the Activity Exemplar lab notebook, you’ll find references to different courses, such as Course 5 (regression modeling and analysis) and Course 6 (machine learning models). However, you are not required to use material from every course. You have the freedom to choose which approach to use and subsequently to choose which course’s material to review for reference.*

Your Python notebook should:

* Include the correct code to build and evaluate a predictive model; to do so, you may choose to use one or more of the following: regression model(s), Naive Bayes, and tree-based machine learning models such as decision tree, random forest, and XGBoost
* Clearly communicate the decisions you made through all stages of the project, including EDA and modeling
* Present the results from your model(s)



**Course 7 executive summary**

Compare the exemplar to your completed executive summary. Your responses might differ in some respects from the exemplar, but that is to be expected. What did you do well? Where can you improve? Use your answers to these questions to guide you as you continue to progress through the course.

***Note:*** *The exemplar represents one possible way to complete the executive summary. Yours might differ in certain ways, such as your specific language and visual design. What’s important is that you have an overall understanding of the purpose and organization of executive summaries for data projects.*

Your executive summary should:

* Include key information that you want to share with your stakeholders
* Use clear and concise language to effectively communicate your results