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.
Be sure to complete this activity before moving on. The next course item will provide you with completed exemplars to compare to your own work. You will not be able to access the exemplars until you have completed this activity.
Scenario
Review the scenario below. Then complete the step-by-step instructions.

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*.

Link to templates:

Course 7 PACE strategy document

Executive summary templates

OR

If you don't have a Google account, you can download the templates directly from the attachments below:

Activity Template_ Course 7 PACE strategy document
DOCX File

Activity Templates_ Executive summaries
PPTX File

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

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

1. Did you complete this activity? 1 / 1 point

- (●) Yes
- ( ) No

Correct
Thank you for completing this activity! Using Python to build and evaluate different statistical and machine learning models is a powerful tool for data analysis. Further, effectively communicating your findings with an executive summary is an important skill for any data professional. Go to the next course item to compare your work to completed exemplars.