<https://olympus.mygreatlearning.com/courses/128824/modules/items/6728030>

Description

**Problem Statement:**

An automobile dealership in Los Vegas specializes in selling luxury and non-luxury vehicles. They cater to diverse customer preferences with varying vehicle specifications, such as mileage, engine capacity, and seating capacity. However, the dealership faces significant challenges in maintaining consistency and efficiency across its pricing strategy due to reliance on manual processes and disconnected systems. Pricing evaluations are prone to errors, updates are delayed, and scaling operations are difficult as demand grows. These inefficiencies impact revenue and customer trust. Recognizing the need for a reliable and scalable solution, the dealership is seeking to implement a unified system that ensures seamless integration of data-driven pricing decisions, adaptability to changing market conditions, and operational efficiency.

**Objective:**

The dealership has hired you as an MLOps Engineer to design and implement an MLOps pipeline that automates the pricing workflow. This pipeline will encompass data cleaning, preprocessing, transformation, model building, training, evaluation, and registration with CI/CD capabilities to ensure continuous integration and delivery. Your role is to overcome challenges such as integrating disparate data sources, maintaining consistent model performance, and enabling scalable, automated updates to meet evolving business needs. The expected outcomes are a robust, automated system that improves pricing accuracy, operational efficiency, and scalability, driving increased profitability and customer satisfaction.

**Data Description:**

* **Segment**: Describes the category of the vehicle, indicating whether it is a luxury or non-luxury segment.
* **Kilometers\_Driven**: The total number of kilometers the vehicle has been driven.
* **Mileage**: The fuel efficiency of the vehicle, measured in kilometers per liter (km/l).
* **Engine**: The engine capacity of the vehicle, measured in cubic centimeters (cc).
* **Power**: The power of the vehicle's engine, measured in brake horsepower (BHP).
* **Seats**: The number of seats in the vehicle, can influence the vehicle's classification, usage, and pricing based on customer needs.
* **Price**: The price of the vehicle, listed in lakhs (units of 100,000), represents the cost to the consumer for purchasing the vehicle.

**Submission Guidelines**

1. There are two ways to work on this project:

**i. Full-code way:**The full code way is to write the solution code from scratch and only submit a final Python notebook with all the insights and observations.

**ii. Low-code way**. The low-code way is to use an existing template notebook to build the solution and then submit a business presentation with insights and recommendations.

The primary purpose of providing these two options is to allow learners to opt for the approach that aligns with their learning aspirations and outcomes. The below table elaborates on these two options.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Submission type** | **Who should choose** | **What is the samebetweenthe two?** | **What is different between the two?** | **Final submission file [IMP]** | **Submission Format** |
| Full-code | Learners who aspire to be in hands-on coding roles in the future focussed on building solution codes from scratch | Perform exploratory data analysis to identify insights and recommendations for the problem | Focus on code writing: 10 - 20% grading on the quality of the final code submitted | Solution notebook from the full-code template submitted in .html format | .html |
| Low-code | Learners who aspire to be in managerial roles in the future focus on solution review, interpretation, recommendations, and communicating with business | Focus on business presentation: 10 - 20% grading on the quality of the final business presentation submitted | Business presentation in .pdf format with problem definition, insights, and recommendations | .pdf |  |

Please follow the below steps to complete the assessment. Kindly note that if you submit a presentation, ONLY the presentation will be evaluated. **Please make sure that all the sections mentioned in the rubric have been covered in your submission.**

**i. Full-code version**

* Download the full-code version of the learner files required for the project.
* Follow the instructions provided in the files to complete the project.
* Write down insights and recommendations for the business problems in the comments.
* Add the link to the GitHub repository containing the GitHub Actions workflow to a markdown cell at the top of the notebook.
  + **Ensure the GitHub repo is a public one.**
* Submit only the solution notebook prepared [format: .html]

**ii.** **Low-code version**

* Download the low-code version of the learner files.
* Follow the instructions provided in the files to complete the project.
* Prepare a business presentation with relevant information and screenshots of the GitHub Actions workflow as well as the insights and recommendations for the business.
* Add the link to the GitHub repository containing the GitHub Actions workflow to a slide in the final submission.
  + **Ensure the GitHub repo is a public one.**
* Submit only the presentation prepared [format: .pdf]

2. Any assignment found copied/plagiarized with other submissions will not be graded and awarded zero marks.

3. Please ensure timely submission as any submission post-deadline will not be accepted for evaluation.

4. Submission will not be evaluated

* if it is submitted post-deadline, or
* if more than 1 file is submitted.

**Best Practices for Full-code Submissions**

* The final notebook should be well-documented, with inline comments explaining the functionality of code and markdown cells containing comments on the observations and insights.
* The notebook should be run from start to finish sequentially before submission.
* It is important to remove all warnings and errors before submission.
* The notebook should be submitted as an HTML file (.html) and NOT as a notebook file (.ipynb).
* Please refer to the FAQ page for common project-related queries.

**Best Practices for Low-code Submissions**

* The presentation should be made keeping in mind that the audience will be the Data Science and AI lead of a company.
* The key points in the presentation should be the following:
  + Business Overview of the problem and solution approach
  + Summary of steps followed and content of the GitHub Actions workflow
  + Key findings and insights that can drive business decisions
  + Business recommendations
  + Focus on explaining the key takeaways in an easy-to-understand manner.
  + The inclusion of the potential benefits of implementing the solution will give you the edge.
* Copying and pasting from the notebook is not a good idea, and it is better to avoid showing codes unless they are the focal point of your presentation.
* The presentation should be submitted as a PDF file (.pdf) and NOT as a .pptx file.
* Please refer to the FAQ page for common project-related queries.

**Power Ahead!**

Rubric

CriteriaBuilding an end-to-end MLOps Pipeline in AzureML Studio- Create a MLOps pipeline in AzureML Studio containing the following components - data preprocessing - model training - model tuning and logging using MLflow - model registration - Register all the above components into a single pipeline - Execute the pipeline

Points17

CriteriaCreate a GitHub Actions Workflow with the necessary Python scripts and YAML files- Download and modify (if needed) the relevant Python scripts (.py files) created previously for the different components of the MLOps pipeline - Create the necessary YAML (.yml) files required for the GitHub Actions worflow - Put the Python scripts and YAML files in a hierarchical folder structure

Points18

CriteriaExecute the GitHub Actions Workflow- Set up the necessary credentials to connect Azure and GitHub Actions workflow - Execute the GitHub Actions workflow

Points5

CriteriaValidate CICD Implementation- Update one of the Python scripts (either prep.py or train.py) - Commit the changes to the GitHub repo to validate CICD implementation

Points5

CriteriaSample Output- A screenshot of the GitHub Actions Workflow - A screenshot showing the complete pipeline execution from AzureML StudioPoints5

CriteriaActionable Insights and Recommendations- Conclude with the key takeaways (actionable insights and recommendations) for the businessPoints2

CriteriaPresentation / Notebook - Overall Quality- Structure and flow - Crispness - Visual appeal - Github Folder Structure OR - Structure and flow - Well commented code

Points8