Google

Reading: Incorporate your project into a portfolio

Throughout this certification you have utilized the PACE framework to develop a portfolio project showcasing your workplace and technical skills. Outside of showcasing your projects on your resume, it is important to have an online portfolio that houses your projects. Unlike your resume, which will change frequently based on the job description of the role you are applying for, an online portfolio will only change each time you complete a project. This reading will guide you through how to display your technical skills and projects on your resume and how to create an online project portfolio.

Add your technical skills and portfolio to your resume

When applying for a data role, your first introduction to a potential employer is through your resume. When crafting your resume, it is important to add the technical skills you acquired from the certificate and your portfolio project to the "Skills'' section of your resume. Be sure to create a resume that incorporates the technical skills listed in the specific job description. Below is an example of how you can add these technical skills to your resume:

Example 1: Adding technical skills to your resume

The following is an example of a "Technical Skills" section that you can include on your resume.

Technical Skills

- Programming Languages: Python
- Python Packages: numpy, Pandas, Scipy, seaborn, Matplotlib, statsmodels, scikit-learn
- Machine Learning Models: regression (linear, logistic), Naive Bayes, decision trees, random forest, AdaBoost, XGBoost

After your technical skills section, you can add a section entitled "Technical Projects", "Data Projects," or "Machine Learning Projects." The title of the section is up to you and depends on the wording of the job description. If the job description repeats the phrase "machine learning," then you may want to title the section "Machine Learning Projects". If you have a mix of data projects that include dashboards, data analysis, modeling, etc. you may want to title the section "Data Analytics Projects" and use bullet points to list the different types of projects. Below is an example of how to list a project on your resume:

Example 2: Add a technical project to your resume

The following is an example of a "Data Analytics Projects" section that you can include on your resume.

Data Analytics Projects

- Classification of TikTok videos: Used statsmodels and scikit-learn to predict whether videos presented claims or opinions to improve triaging process of videos for human review
- Classification of Waze data: Built decision tree, random forest, and XGBoost to predict Waze user churn
- Used multiple regression to predict taxi fares, data that would be used as part of a suite of models to optimize revenue for the New York Taxi and Limousine Commission and its drivers

The technical skills and projects you add to your resume should also be reflected in your online project portfolio. As you learn more programming languages, technical packages, and models, you need to add them to your resume. The same technical skills you add to your resume should be showcased within the projects you have in your online portfolio. The next section highlights where to build an online portfolio and what to include in your portfolio.

Where to create your online portfolio

There are many platforms that data professionals use to host their online portfolios. The first step is to choose a platform that suits the type of projects you want to showcase. Google Sites is good for blog-style portfolios. GitHub and Kaggle are better for hosting code-based portfolios. Tableau is great for sharing your visualizations. Create an account on the platform that you chose.The following links have steps that explain how to set up accounts on various platforms:

- Set up an account on GitHub
- Set up an account on Kaggle
- Set up an account on Tableau Public
- Set up a site on Google Sites

Since Github is one of the most popular platforms for creating an online project portfolio, the next section will demonstrate what to add to your GitHub Portfolio in more detail.

What to add to your GitHub Portfolio

After setting up a github account, you will need to create separate repositories for each individual project. Each repository will contain all of your project files and a README.md file. A README is a markdown-based text file that provides an overview of your project. The following sections are great to include in your README:

- Project Title
    - Including a descriptive title that states the type of analysis and project draws in your prospective employer. Do not title your project “Portfolio Project”; instead try adding the modeling algorithm and data used to your title. For example: “Natural Language Processing of Election Day Tweets.”
- Project Overview
    - A project overview should be a few sentences long stating the problem you solved, what data was used in the project, and your modeling results.

- Business Understanding
    - You should have a section that showcases the stakeholder(s) and the business problem you tried to solve. Feel free to add citations of research you did on your business problem here as well.
- Data Understanding
    - Explain what data you used in your analysis, the timeframe of the data, and any data limitations. This is also a good section to add visualizations of your exploratory data analysis.
- Modeling and Evaluation
    - This section should detail what models you used and the corresponding evaluation metrics.
- Conclusion
    - In the conclusion section explain the recommendations you have in solving the business problem and highlight any future steps you will take to expand on your project.

For more information on how to craft README files, checkout GitHub's "About READMEs" article.

The Github repository README (shown below) uses the New York City Taxi & Limousine Commission data that you have seen in your portfolio project throughout the program. This example expands beyond what was given in your original project description by including domain knowledge under the "Business Understanding" section. Domain knowledge demonstrates to a prospective employer your ability to do research before conducting a technical analysis. It is important to create a clear, concise README that summarizes your business understanding and technical findings.

Example Project: README

Predicting Taxi Gratuities in New York City

Overview

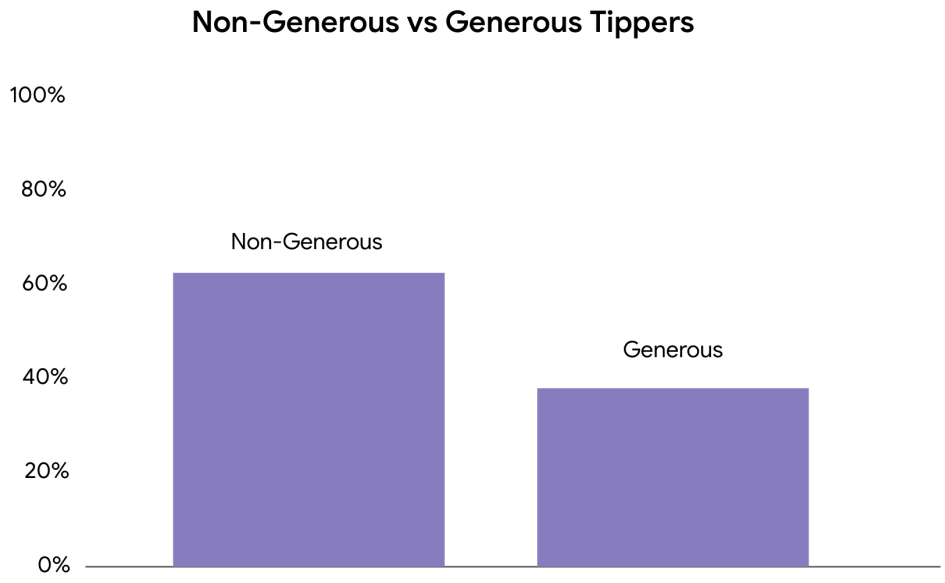
The goal of this project was to create a multiple linear regression and random forest model to predict high rider gratuity or not. This project utilized yellow taxi trips taken in New York City during 2017. The final random forest model performed with 86% accuracy and 72% precision determining what features were most important in separating low tippers from high tippers. Based on the model, the duration, distance, and cost of the trip were most influential in determining a generous tipper (>20%) vs a non-generous one (<20%).

Business Understanding

According to salary.com the average salary for a New York Taxi Driver is around $45,000. This salary is significantly low compared to a median rent value of $6,500 per month. It is important to understand what factors encourage riders to leave tips in order to help drivers obtain a livable wage.

Data Understanding

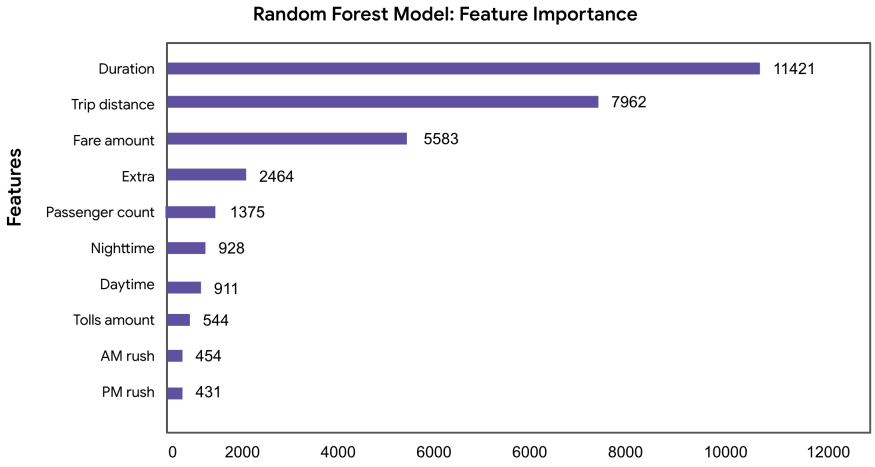
The NYC Taxi and Limousine Commission data came from NYC.gov. The data consisted of approximately 408k unique trips and 18 features. The features included information on trip duration and destination, vendor used, toll information, and payment type. The bar chart below shows the breakdown of how many generous tippers (>20%) versus non-generous tippers that exist in the data set.



In connection to this, a feature was engineered to represent if a ride was taken during rush hour or not. Multiple redundant columns were dropped and reformatted into the proper data type.

Modeling and Evaluation

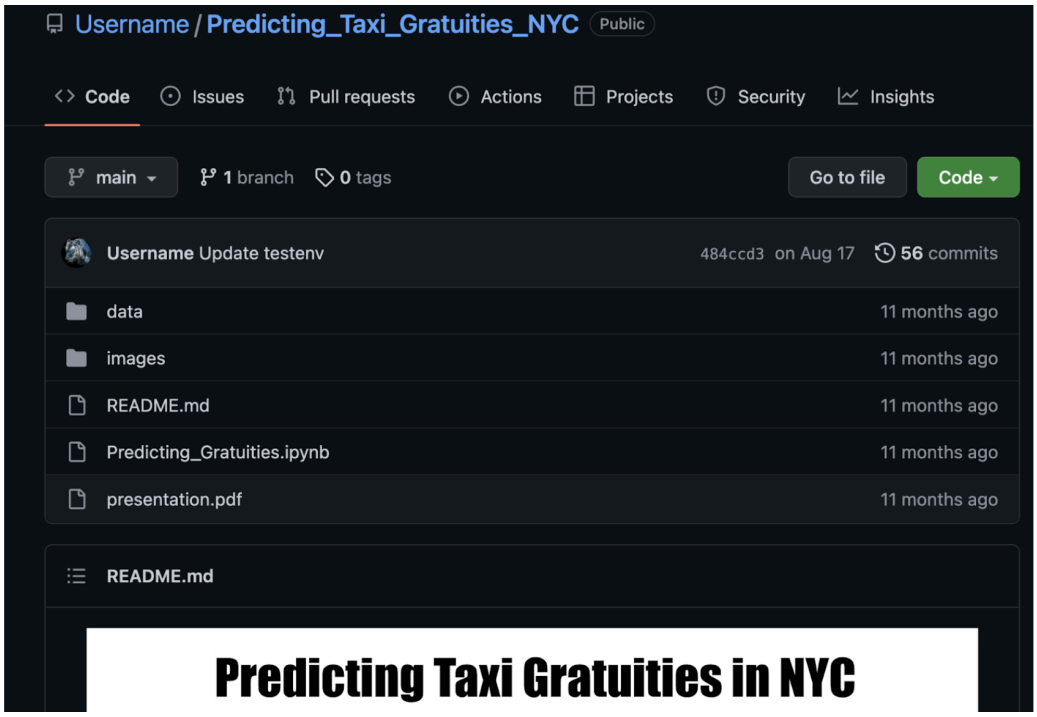
A random forest model comprising 100 decision trees was used to determine feature importance in who would tip generously or not. The below plot shows that trip duration, distance, and the cost of a fare were the Top 3 most important factors in determining a generous tipper from a non-generous one. The overall model performed with 86% accuracy and 72% precision.



Conclusion

This model can benefit Taxi Drivers in knowing if they will be tipped generously or not; however, running a parametric model to determine how much each variable will influence the actual price of the tip. In the future, adding more information on a rider's past tipping behavior may also be beneficial in helping the stakeholder address their business problem.

Outside of the README file, it is important to have the data you used, cleaned up Python notebook files, a presentation, and any images you may have used on your GitHub repository.



Checkout this additional resource from DataQuest that walks you through how to add files to your online Github portfolio. The goal is to have all project information in one repository that will help an employer understand your project, run your code, and clearly know your business recommendations.

Key Takeaways

- You should review the job description's technical skills and add the applicable skills to your resume to increase your chances for being called for an interview.
- Having your data projects on your resume is a great way to showcase your hands-on technical experience for various data roles.
- Github is a great online platform for building an online portfolio of coding projects that can be seen by any prospective employer. Keep in mind that there are other platforms: Kaggle, Google Sites, Tableau Public, Medium, and more R can showcase your technical writing, data visualization, and coding skills.
