CAREERFOUNDRY

Achievement 5 Project Brief: Data Ethics and Applied Analytics

Introduction

Congratulations! You've been hired as a data analyst by a well-known global bank. Your job is to provide analytical support to its anti-money-laundering compliance department. This will involve a variety of data-related projects that help the bank assess client risk and transaction risk, as well as reporting on metrics. You'll also help build and optimize models that assist the bank in running their compliance program more efficiently. Many challenges await you that will test your technical skills and ability to handle data-related ethical dilemmas.

Objectives

This Achievement will introduce a variety of new concepts and processes, many of which fall under the realm of advanced data analytics. Don't worry if you don't get everything right away, as you'll revisit some of these more complex topics in Achievement 6.

Below are just a few of the topics you'll encounter:

- The characteristics of big data, how data analysts use big data, and the challenges of extracting knowledge from big data
- The impact of data bias and ethics on how data is used, shared, collected, and protected
- The fundamentals of data mining, including techniques for data mining and how it drives decision-making
- Predictive analysis and models such as linear regression
- Time-series analysis and time-series forecasting
- The basics of GitHub and how you can use it to refine your skills, collaborate with colleagues, and display a portfolio of work
- What to include in your portfolio when applying for jobs

Context

Data analytics is an exciting profession if you're curious, enjoy problem solving, and want to make a difference. Indeed, the satisfaction of being able to measure your impact and drive decision-making is what attracts many to the profession. The amount of data that's being analyzed is minuscule compared to the amount of data that's being produced. We've not even begun to scratch the surface of big data, which is growing rapidly. While machine learning can certainly help in this area, vast amounts of knowledge and insights are still waiting to be found by those with the right skill set and know-how.

Technology and an interconnected world are vastly increasing the amount of data collected. Everything from how much it rained in Seattle last year to how many text messages were sent worldwide and the most-watched movies on Netflix all become part of big data. However, the ways in which data is collected, used, and shared can be harmful to both individuals and society. Data collected on individuals in particular comes with responsibility. For these reasons, the data analyst should be guided by a strong ethical foundation and be able to discuss ethical concerns with their coworkers and employer.

Besides being aware of data ethics and knowing how to raise ethical concerns with stakeholders, the data analyst needs to know how to derive useful information from big data. This is where data mining, predictive analytics, and time-series analysis and forecasting come into play.

Finally, the data analyst needs to understand GitHub. Not only is GitHub the standard for any data analyst portfolio, it's also a great way for analysts to collaborate on projects and learn from others in the industry. What's more, displaying your SQL and Python skills in GitHub will impress future employers and serve as proof that you have the skills they require when applying for jobs.

Project Tasks & Deliverables

Throughout this Achievement, you'll work on your project from one Exercise to the next, completing tasks as you go. For each task, you'll submit a deliverable that makes up a piece of your project. Below is a breakdown of your tasks and deliverables by Exercise:

Exercise 1: Intro to Big Data

- Describe the characteristics of structured and unstructured data
- Identify the applications and limitations of big data
- Research software tools for handling big data

Exercise 2: Data Ethics: Data Bias

- Identify the types of bias encountered in the workplace
- Suggest ways of controlling for bias and communicating concerns to stakeholders

Exercise 3: Data Ethics: Security & Privacy

- Identify ethical dilemmas in the workplace
- Suggest ways to raise ethical concerns with stakeholders
- Research ethical dilemmas and security and privacy laws in different countries

Exercise 4: Intro to Data Mining

- Carry out steps in the data mining process, including data cleaning and descriptive statistics
- Create a decision tree model to test the outcomes of an analysis

Exercise 5: Intro to Predictive Analysis

- Differentiate between logistic and linear regressions models
- Analyze the output of a linear regression
- Identify the correct predictive model for different scenarios

Exercise 6: Time Series Analysis & Forecasting

- Create and describe the characteristics of a time series
- Create a simple moving average in Excel and describe its characteristics
- Differentiate between stationary and non-stationary time series
- Research different time forecasting models

Exercise 7: Using GitHub as an Analyst

- Create a GitHub account and repositories
- Host your SQL and Python work from Achievement 3 and 4 in GitHub

Exercise 8: Preparing a Data Analytics Portfolio

- Curate your projects from the Data Immersion Program in advance of your Job Prep course
- Start creating a storyline around your projects