AIP Project Case Study Orientation

Vocational Learning Outcomes (VLOs) Covered in this AIP Project Case Study

- Collect, house, extract, manipulate, maintain, and mine data sets that respond to organizational, financial, or market needs.
- Recommend different systems and network architectures, artificial intelligence, and data storage technologies to support data analytics and Big Data.
- Design and apply data models that meet the needs of a specific operational process or business model.
- Develop software applications, algorithms, and artificial intelligence models to manipulate, correlate and reduce data sets and produce project documentation and reports.
- Collaborate effectively with diverse teams to design and present data visualizations that communicate Big Data concepts and information to stakeholders and business professionals.
- Apply business analytics, business intelligence tools and research to support evidence-based decision-making that helps an organization meet their stated objectives.
- Identify and assess data analytics and Big Data business strategies and workflows to respond to new opportunities or provide project solutions.
- Implement data security solutions in compliance with corporate security policies, ethical standards, and industry regulations.
- Deliver data-oriented projects using data science, business analysis, and project management principles, tools, and techniques.
- Develop artificial intelligence solutions to support administration, decision-making, planning, risk management, logistics, manufacturing, smart devices, and robotics.

Essential Employability Skills (EESs) Covered in this AIP Project Case Study

- Communication
  - It helps to communicate clearly, correctly, and concisely in different forms. These include oral, written, and visual.
- Numeracy
  - This skill set helps to solve mathematical operations effectively with accurate precision.
- Critical thinking & problem solving
  - It is a systemic approach to attempting to resolve problems by analyzing the pros and cons of a decision.
- Information management
  - It helps to locate, select, organize, and document information with the use of technology by analyzing aspects and gathering information from a variety of sources.
- Interpersonal Skills
  - This skill set is important as it helps to respect others' opinions or input. It helps to build teams and maintain relationships to achieve overall team or organizational goals.
- Personal Skills
  - These soft skills are important in developing employability talents, such as dependability, adaptability, and problem-solving skills.

Week 1

Applicable VLOs or EESs for This Week's Case Study

None for this week.

This Week's Detailed Case Study Information

Congratulations you have been hired as an ML Engineer for Flysafe Analytics located in Mississauga ON.

Flysafe Analytics is a sub-division of Flysafe Airlines which is an international Airline company. All the data management and processes for Flysafe Airlines are handled by Flysafe Analytics. The analytics company does sales forecasting, social media analysis and digital marketing analysis. It has a strong team of Data Scientists, Data Engineers, ML Engineers, MLOps and Business consultants with in-depth knowledge and experience in solving data-related problems such as social media analytics, customer insights, customer analytics, forecasting, sales and promotion planning, scheduling, pricing and revenue management.

The customer can book and buy tickets from the website and can leave reviews for particular flights regarding flights and customer service experiences. The website has been storing data for a few years now. Since there are lots of flights on daily basis, flight reviews are there as well. A few years back they used to have an on-premise infrastructure to store all these data but now they have migrated to the cloud and are storing data in dynamodb

In the 21st century, data has become an integral part of almost every aspect of our lives. From the way we shop and communicate, to the way we work and make decisions, data plays a critical role in helping us navigate an increasingly complex and interconnected world.The importance of data insights in the 21st century cannot be overstated. By analyzing and interpreting data, businesses and organizations can gain valuable insights into their customers, operations, and markets. This, in turn, allows them to make more informed decisions, optimize their processes, and improve their performance.

The sales department has reported that ticket sales have fluctuated drastically and as a part of finding out the reason for changes in sales, the data team is analysing data from different domains. One of the team is analysing the customer review collected from the website. Data from earlier years were stored in an on-premise system and recent data are stored in DynamoDb. You are assigned to the same team as ML Engineer. The objective of this team is to find out the sentiment from the customer review. Since the data is present on both on-premise and cloud your team's responsibility is to find the best ways to collect the data and perform sentiment analysis in it.

Your Analytics team consist of

- Jim Cooper (Lead Data Scientist)
- Leonard Galecki (Data Scientist)
- Howard Helberg (Data manager)
- Amy Bialik (Project Manager)
- Kunal Koothrappali (Data Engineer)

You will be working closely as an ML Engineer with these team members.

Meet Jim, the highly knowledgeable and experienced data scientist who has been working in the field of natural language processing (NLP) for a decade. Throughout his career, he has contributed his expert skills to various NLP projects and even dabbled in voice recognition research before joining Flysafe Analytics. Next, meet Amy, the well-rounded project manager with a diverse background in managing a variety of projects, including software development and data analytics. Currently, she is using her skills to successfully oversee data analytics projects at Flysafe Analytics. Last but not least, we have Kunal, the skilled data engineer proficient in python, excel, and SQL. He is responsible for ensuring smooth data pipelines and preparing data for predictive and prescriptive modeling.

Week 1 Onboarding Expectations and Participation

Your task this week is to participate in training and orientation for Flysafe Analytics. You will participate in a variety of exercises that are designed to get to know you better and understand your role within the team. You will participate in Team-building exercises that prepare you for success within the Project. As with any position, you will have an excellent opportunity to build on your skills as a leader so long as you put forth your best effort. Use this week to develop a communication plan with your team and be ready to dive into the deliverables starting next week.

Note: You can make any assumptions that are deemed necessary for each case on a week-by-week basis. You will not be provided direct answers or 100% of the information necessary to complete each deliverable. Instead, focus on delivering the highest quality outcome possible to highlight your talent as a group. You would be presenting these deliverables to Amy and would want to ensure that the work is of the highest quality.

This section will be available to you for the entirety of the project. However, each subsequent week's case study information may only be available for that week. Be sure to download and save this week's information for future use.

Week 2

Applicable VLOs or EESs for This Week's Case Study

- Collect, house, extract, manipulate, maintain and mine data sets that respond to organizational, financial, or market needs.
- Recommend different systems and network architectures, artificial intelligence, and data storage technologies to support data analytics and Big Data.
- Identify and assess data analytics and Big Data business strategies and workflows to respond to new opportunities or provide project solutions.

This Week's Detailed Case Study Information

There is a meeting coming up with Jim, Howard, Amy, and you to discuss the tools and technology stack that will be used for the current project. The selection of these tools and stack will depend on the nature of the problem at hand, the skillset of the team, and the budget allocated for the project. It is always helpful to do a thorough market analysis and have a good understanding of the issues before attending the meeting. There are many tools that can potentially be leveraged to provide a solution for this project, so it's important to conduct a thorough analysis and come up with the ideal stack for the specific use case.
After the meeting with the team, a decision will be made on the tools and technology to be used for the project. As an ML engineer, it's important to stay up to date on emerging technologies, their alternatives, and their limitations.

Having domain knowledge about the data is also crucial for understanding the context of the problems you are trying to solve. In this case, having an understanding of industry jargon and terminology related to the airline and customer service industries will help you better understand the sentiments of customers.

It's also worth noting that staying up to date on the latest developments and best practices in ML and related fields will be important for the success of this project. Additionally, having strong communication skills and the ability to work effectively in a team will be key for success.

Deliverables for This Week's Case Study

1. Research and gain a solid understanding of the basic principles and mechanisms behind how machine learning (ML) works.
2. Prepare a brief report that covers the terminology and jargon commonly used in the airline industry.
3. Create a block diagram that illustrates the IT infrastructure needed for this system.
4. Identify and list the various tools, services, and software that will be required to integrate all of the components and make the system functional.

Week 3

Applicable VLOs or EESs for This Week's Case Study

- Develop software applications, algorithms, and artificial intelligence models to manipulate, correlate and reduce data sets and produce project documentation and reports.
- Design and apply data models that meet the needs of a specific operational process or business model

This Week's Detailed Case Study Information

Data migration refers to the process of transferring data from one system to another. While this may seem straightforward, it actually involves changing the storage location and potentially the database or application as well. In the context of the extract/transform/load (ETL) process, any data migration will include at least the transform and load steps, which involve preparing the extracted data and loading it into the target location.

Flysafe Analytics used to store its data on-premise infrastructure, but recently started using cloud solutions. The old data has not yet been migrated to the cloud, and the company is planning to do so. Jim believes that now is the right time to migrate the old data to the cloud so that it can be used for the sentiment analysis project.

Before beginning any transformation and analytical processes, it's important to define the scope of the project and document it. This will give the team a clear understanding of the data requirements and help them stay on track.

To ensure that the project stays on schedule, the team will be using agile methodologies and will have either weekly or biweekly deliverables to submit. Feedback from each sprint will be documented to help plan for the next one. Daily stand-up meetings will be held with all team members to discuss tasks and any issues that arise, and to find solutions. These meetings will be brief to allow the team to focus on their daily work. If any issues come up that could potentially delay task completion, it's important to inform Amy as soon as possible so that she can reassess the situation and adjust the plan accordingly. Finally, the scope of the project, including deadlines, milestones, and deliverables, will be finalized in a meeting with Jim, Amy, and Howard once the entire team agrees on the timeline for each task and deliverable. It's worth noting that data migration is also a part of the project.

- ***Purpose***

1. What is the purpose of doing sentiment Analysis?

- ***Business objectives***

  Prepare a business case for your project. What business objectives are expected at the end of this project?

- ***Technical objectives***

  Based on the business objectives, prepare project technical objectives and map them with business objectives. Filter out the technical aspects which are recommended by your team to achieve business objectives.

- ***Project Description – requirements, scope, deliverables, stakeholders***

1. Define the scope of the Project.
2. Create a **Preliminary Scope Statement.** As part of any project, be it in Business, Project Management, IT, Biotechnology, etc., you must build a preliminary scope statement. The scope defines the boundaries of the project, what you will be looking at closer, and what you hope to accomplish.
3. The Preliminary scope statement helps to focus your efforts and research and the point of this week is to create it, then discuss it and refocus it. In the workplace, groups will often change and evolve their scope as they progress through a project, it is important that your research project reflects this as well.
4. Define the requirements of the projects with the deliverable and stakeholders.

Deliverables for This Week's Case Study

Project Charter

1. Elaboration on the purpose of implementing sentiment analysis application, how it works and is implemented.
2. What is supposed to be the budget of the project?
3. Outline all aspects of a project, including all related activities, resources, timelines, and deliverables, as well as the project's boundaries.
4. Write **business objectives** with which Flysafe Analytics can achieve specified goals, increase productivity and decrease human labour. Write ways to increase the efficiency of the project. Apart from that how proposed technological tools can be used from a security point of view?

Technical Deliverables:

1. Make sure all the required applications are installed and configured to implement this use case.
2. Create an AWS account.
3. Identify the different file formats to be used for this use case
4. Get a sample data file from HDFS to your local machine
5. Write a python script to convert the parquet file to CSV and vice versa and test it.
6. Try out commands used for reading and writing data to HDFS

Week 4

Applicable VLOs or EESs for This Week's Case Study

- Develop software applications, algorithms, and artificial intelligence models to manipulate, correlate and reduce data sets and produce project documentation and reports.

This Week's Detailed Case Study Information

Amy wants you to identify the triple constraints for the project (Time, scope and cost). Before doing that, you need to have a clear picture of the task involved in the project. First, identify the major milestones of the project and allocate time to each milestone. Since it's a machine-learning problem, you need to take care of each step in the ML workflow and allocate time for it. As mentioned in earlier weeks, an agile team including Jim, Amy, Howard, Kunal and you would be the human resource for the project. The project will last for about four months. Since you have the estimates of the resource and time you can prepare the budget for the project. Once you are done with the timeline estimate the budget for the project.

Project planning should accommodate all these steps. Divide the Sentiment analysis projects into these steps and create an action plan.

1. Identify goals/priorities/initiatives.
   - Short term
   - Identify goals/priorities/initiatives.
   - Identify what, when, and who; and are measurable
2. Prepare a tentative schedule for implementation at the client site (table format). A GANTT chart may prove useful to highlight the breakdown of your timeline and work strategies
3. Budget & resource allocation (table format): personnel, software, hardware, etc. Include names, descriptions/specs, unit price, quantity required, total item cost, and total project cost to the client.

Deliverables for This Week's Case Study

1. Action Plan - progress-to-date
   - Estimate the total budget of the project that Flysafe Analytics must pay
   - Make a SWOT analysis report that shows the strength, weaknesses, and threats faced by building a Sentiment Analysis model and associated technologies
2. Determine the short-term goals and transform the strategies you make in the last week into a specific performance target.

Week 5

Applicable VLOs or EESs for This Week's Case Study

- Develop software applications, algorithms, and artificial intelligence models to manipulate, correlate and reduce data sets and produce project documentation and reports.

This Week's Detailed Case Study Information

Cloud computing is one of the hottest catchphrases in business today. It has transformed the way organizations store, access and share information, collaborate and manage computing resources. With the advent of the internet, cloud computing has provided new ways of conducting business by allowing companies to rise above the conventional on-premises IT infrastructure.

Amy has asked you to find the ideal service to store the data in the cloud. There are many cloud service providers offering different kinds of storage and computing services in the cloud. The ideal service will be able to store the data and would be cost-efficient. Try to understand the use case and find out how costing will be calculated for chosen cloud services. Since DyanmoDb is used as a Database, choosing some service within the AWS domain would be a wise decision. Be ready to explain the reason behind the choice of services. You can always refer to the documentation provided by the cloud service providers to learn more about them and justify your use case.

Before shipping data to a different location or performing any kind of transformation, the metadata(data about data) should be clear. Without knowing the metadata or structure of the data, no operation should be performed. To create a table in the DynamoDb you need to know the structure of the data. The data may be zipped or raw. There are different possible file formats and before diving into ETL processes make sure you are clear about the file formats and structure.

- Explore the format of your data and identify the required field
- Analyze the data type required for each field
- Research the use case scenario of using DynamoDb and other NoSQL databases.

Deliverables for This Week's Case Study

1. Identify the column or fields in the data and the data type required for it.
2. Upload the data to AWS S3 and export it to DynamoDB
3. Use the boto library in python to connect with DynamoDB and test the connection
4. Display the first ten rows in DynamoDB with python API.

Week 6 – Mid Term Week

Mid-Term Panel Evaluation Preparation

The team will prepare for the Mid-Term Panel Evaluation this week. For the Team Presentation create a professional multimedia presentation highlighting the key aspects of your project thus far. Please see Moodle for full details.

Presentation and Oral Delivery

CONTENT

- Overview of work in the Flysafe Analytics
- Highlight three key areas you find of interest:
  - Two areas related to weekly work completed
  - One area to highlight PD or other activity
- Apply reflecting skills
- Present the importance/Benefit of work to Flysafe Analytics.

Week 7

Applicable VLOs or EESs for This Week's Case Study

- Identify and assess data analytics and Big Data business strategies and workflows to respond to new opportunities or provide project solutions.
- Implement data security solutions in compliance with corporate security policies, ethical standards, and industry regulations

This Week's Detailed Case Study Information

Assessing the risk of a project is an essential part of the planning process. It involves identifying potential hazards and determining the likelihood and impact of those risks occurring. At Flysafe Analytics, there is a strict policy of providing data on a "need to know" basis in order to reduce the risk of data privacy violations.

As part of this policy, Amy has asked you to carefully analyze the datasets and remove any unnecessary fields before providing them to the analytics team. The reviews that you will be working with have been collected from a website and may contain personally identifiable information (PII). However, the use case for this project does not require this information, so it's important to make sure that you remove it before proceeding with any data analytics processes. This will help protect the privacy of individuals and reduce the risk of data breaches.

Risk Analysis

1. Risk identification and mitigation strategies
   This week, your team will identify the potential risks associated with your solutions or idea. Using the Risk matrix conduct risk analysis and find the possible risks faced by the company because of performing sentiment analysis and the weakness pointed out earlier in weeks 4 using SWOT Analysis what are the required strategies to mitigate them?
2. Software/hardware methods, techniques and tools needed.
   Moreover, based on the above research provide Software/hardware methods, techniques and tools that might be required for risk mitigation.

Deliverables for This Week's Case Study

1. Create a tabular detailed work highlighting risk analysis using Risk Matrix indicating risks, and vulnerabilities as per your solution or hardware used.
2. Identify any security issues.
3. Propose security tools to prevent any security-related issues.
4. Write a short report on how PII can be tied to security threats.

Week 8

Applicable VLOs or EESs for This Week's Case Study

- Develop artificial intelligence solutions to support administration, decision-making, planning, risk management, logistics, manufacturing, smart devices, and robotics.

This Week's Detailed Case Study Information

To perform supervised learning, it's necessary to have a labeled dataset. In this case, the labeled dataset consists of reviews with their corresponding sentiments labeled as either positive or negative. Fortunately, Flysafe Analytics has outsourced the task of labeling the data, so a labeled dataset is now available for training and testing purposes. The next step is to initiate the machine learning (ML) workflow.

Jim has asked you to build a model that can analyze the sentiments of reviews and classify them as either positive or negative. He is skeptical of your skills in building models for AI/ML workflows, so it's up to you to prove him wrong. To do this, you will need to follow all the steps of the ML workflow and build a model that can accurately identify the sentiments of reviews. Once the model is built, you can use a confusion matrix to evaluate its performance and discuss the results with Jim. By successfully completing these tasks, you can demonstrate your capabilities in building models for ML workflows and potentially gain Jim's confidence in your skills.

Once you have collected your dataset, you will need to preprocess the text data by removing any irrelevant or redundant information, such as HTML tags or special characters. You may also want to consider stemming or lemmatizing the words in the reviews to reduce the dimensionality of your dataset.

Deliverables for This Week's Case Study

1. Make a report comparing the cost incurred when using the cloud solution and in-house system for ML purposes for this project.
2. Prepare an architecture diagram for the components used.
3. Identify the different libraries of python to be used for data cleaning, transformation, and modelling processes.
4. Use a confusion matrix to evaluate your model performance and explain its significance in this use case.
5. Build a model using an ideal algorithm for this scenario and rely on the F1 score and accuracy to choose the model.

Week 9

Applicable VLOs or EESs for This Week's Case Study

- Design and apply data models that meet the needs of a specific operational process or business model.
- Collaborate effectively with diverse teams to design and present data visualizations that communicate Big Data concepts and information to stakeholders and business professionals.

This Week's Detailed Case Study Information

Data scientists implement exploratory data analysis tools and techniques to investigate, analyze, and summarize the main characteristics of datasets, often utilizing data visualization methodologies. EDA techniques allow for effective manipulation of data sources, enabling data scientists to find the answers they need by discovering data patterns, spotting anomalies, checking assumptions, or testing a hypothesis.

Flysafe Analytics has a Power BI Premium subscription for its data visualization needs. This week you are collaborating with Leonard to perform EDA on the cleaned data. In earlier weeks you cleaned the data set, now it's time to explore more about the data set and find anomalies and outliers. You will be using both python libraries and Power BI to visualize the data and learn more about it.

This step is a continuation of a model-building process that you started in earlier weeks. Data visualization helps to analyze data, and find hidden patterns, anomalies and outliers.

- Use different plots and graphs to visualize the data
- Compare the visualization built using python libraries and power BI.
- Note down the facts or notions in the dataset you were able to see because of the visualization
- Build the ML model again with the insight you received after EDA with visualization.

Deliverables for This Week's Case Study

1. Perform Exploratory Data Analysis (EDA) on the data and build the model using a classification algorithm.
2. List the insights about data you were able to see due to data visualizations.
3. Build a model after doing EDA with visualizations and compare the models.
4. Discuss the use case scenarios for Data visualizations using python libraries and Power BI.

Week 10

Applicable VLOs or EESs for This Week's Case Study

- Develop software applications, algorithms, and artificial intelligence models to manipulate, correlate and reduce data sets and produce project documentation and reports.
- Deliver data-oriented projects using data science, business analysis, and project management principles, tools, and techniques.

This Week's Detailed Case Study Information

Data analytics is an iterative process. Analyst spending more time exploring the data and building models will develop more knowledge of the data. Collaboration with Data scientists and domain experts will help to gain a unique perspective on data insights.

You have developed a model using an algorithm to classify the transactions in weeks 8 and 9 Use different algorithms and compare the performance of models. The use of an algorithm depends on the use case, not all algorithms will be applicable to all classification problems. You will be discussing the choice of your algorithm with Jim and Leonard so you should get facts and talking points ready.

Flysafe analytics is planning to integrate the developed model into a web application using REST API. At the end of every day, the system would send the batch data of review to the application and identify sentiments. The aim of the system is to track the sentiments of the customer on daily basis. Amy has asked you to come up with the IT infrastructure to make this solution possible.

- Research the cross-fold validation and evaluate your model using the same.
- Use different folds for k-cross-fold validation and analyze the result.
- Research how the built models can be saved and reused.

Deliverables for This Week's Case Study

1. List the pros, cons, and trade-offs of using your choice of algorithm.
2. Make a report highlighting the performance of different models and explaining how cross-validation assists in measuring algorithm performance.
3. Research the ways to implement these models in real-time reviews and predict the possible outcomes.
4. Build the block diagram for the application where the model built can be used in real time.

Week 11

Applicable VLOs or EESs for This Week's Case Study

- Develop software applications, algorithms, and artificial intelligence models to manipulate, correlate and reduce data sets and produce project documentation and reports.
- Deliver data-oriented projects using data science, business analysis, and project management principles, tools, and techniques.

This Week's Detailed Case Study Information

Once teams move from a stage where they are occasionally updating a single model to having multiple frequently updating models in production, a pipeline approach becomes paramount. A machine learning pipeline helps in automating machine learning workflows. Machine learning pipelines are iterative as every step is repeated to continuously improve the accuracy of the model and achieve a successful algorithm. The code is split into more manageable components such as data validation, model training, model evaluation, and re-training triggering.

In the case of normal machine learning workflow, the model is the product and in automated workflows, the pipeline is the product. Even an ad hoc model can be deployed in real time depending on the use case. Save the ad hoc model (without pipeline) using pickle and Joblib libraries

In earlier weeks you have tried to build an ad hoc model and the code might not be reusable it can be mostly used for research purposes and test environments. Building with the same fundamentals Leonard has asked you to build a pipeline for the same workflow, this time making the reusable code. Refer to Sklearn documentation to make pipelines.

To make the work reusable and implement the modular design in ML workflow create pipelines.

- Research about the ML pipelines and the Benefits of using pipelines
- Create functions for every step that you build in ML workflow in earlier weeks
- Create a pipeline for the whole process
- Compare this modular approach with the normal steps that you followed in previous weeks

Deliverables for This Week's Case Study

1. Submit both the pickle and joblib file.
2. Submit the python scripts created for the pipeline implementation.

Week 12

Applicable VLOs or EESs for This Week's Case Study

- Deliver data-oriented projects using data science, business analysis, and project management principles, tools, and techniques.

This Week's Detailed Case Study Information

Training the model from scratch usually involves one or more of these components: making use of a widely accepted sentiment lexicon, scoring sentiment by human experts, labelling data by agency contractors or research assistants, and tuning the model that performs well on the rest of the dataset. This process may be costly and time-consuming. On the other hand, using pre-trained classifiers saves a lot of time. These models are easy-to-use with a couple of lines of code, but the specificity of their training datasets might constrain them.

Since the company is planning to implement the solution to achieve insights on a daily basis, it is mandatory that the team should come up with the ideal solution to solve the problem. Jim and Leonard suggested that pre-trained models for NLP should be tried out too. There are models readily available to use in the market that can predict the sentiment of the given data. Before implementing the models that the company has built, Amy asked you to collaborate with Jim and Leonard to experiment with the pre-trained model.

Some of the pre-trained models available are Vader, Happy Transformer, TextBlob, and Google cloud NL API. Leonard requested you go through the documentation of these models. The implementation of these models is pretty straightforward and simple. Some of them are open source and others are not so you should consider the cost before adopting them.

This week, you need to

- Identify the open source and paid services for the pre-trained model.
- Make a report comparing the cost and time estimates for building your own model and using the pre-trained model.

Deliverables for This Week's Case Study

1. In this week, your team needs to submit:
    - Use any two open-source pre-trained models to predict the sentiment on airline reviews.
    - Compare the accuracy and precision of your model and the pre-trained model.
    - Analyze the complexity and speed of your model and pre-trained model and log it

Week 13

Applicable VLOs or EES for This Week's Case Study

- Identify and assess data analytics and Big Data business strategies and workflows to respond to new opportunities or provide project solutions.
- Deliver data-oriented projects using data science, business analysis, and project management principles, tools, and techniques.

This Week's Detailed Case Study Information

Submission of Project Report + Practice Presentation

- Finish Project Report for submission your final submission is due this week. Be proud of the work you have completed in this project, now you can spend time polishing your presentation and making sure you will capture the stakeholder's attention in a positive way.
- Review APA Guidelines and ensure your project has followed them. This is particularly important.

Hone your presentation skills.

- A Presentation for your Sentiment analysis project is meant to highlight your research findings and the conclusions, opportunities, and best practices that you can be followed on other projects. The analysis of your findings is one of the most important parts and should be conveyed in your presentation.

Deliverables for This Week's Case Study

1. Final Project Report – this is your final document with all supporting resources: including any appendices. Bibliography and reference in APA format required.
2. Feedback Video
    - Prepare to answer questions regarding the project on client expectations, Job Market, and on how you will sell your product

Week 14

Preparing for Your Final Week Activities

It is the end of your work term. Your supervisor is grateful for your efforts. The final week contains activities which include both individual and teamwork efforts. Take this opportunity to shine bright in the final activities.

Final Week Deliverables and Format Requirements

Your supervisor will provide you with more detail about the Final Week responsibilities.

AIP Project Completion

Following completion of the Final Week activities, you will be notified by your supervisor if you pass or fail the AIP Project.

Appendix

Acronym Used

HDFS: Hadoop Distributed File system
PCA: Principal Component Analysis
NLP: Natural Language processing
PD: Personal Development
HQL: Hive query Language
SQL: Structured Query Language
RBAC: Role Based Access Control
ML: Machine Learning
DE: Data Engineer
PM: Project Manager
PII: Personal identifiable information
AI: Artificial Intelligence