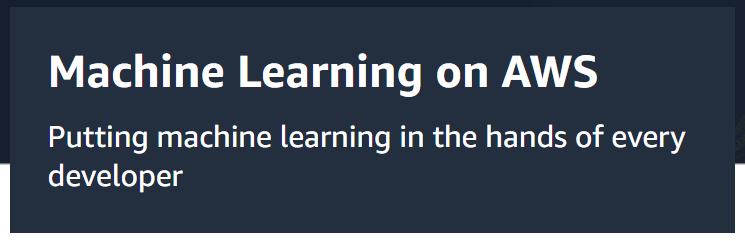


## Project Overview

Because of time difference, I’d like to finish this final project individually.

I want to choose machine learning + cloud service as the topic of my final project. My objective is to deploy some machine learning models (e.g., decision tree, random forest, Naïve Bayes model) on cloud and do data analysis automatically with the help of AWS.

In AWS website, I find a series of cloud services called “AWS Machine Learning services”. In fact, machine learning algorithms usually cost a lot of computing resources to do training and are required to configure specific work environments which are very complicated. From my perspective, AWS cloud services can help solve these problems. In addition, since my major is ADS, it's a good chance for me to practice not only what I learned in IST 615, but also how to combine cloud management and data analysis together.



## Cloud Services:

1. Amazon SageMaker : Amazon SageMaker - Build, train, and deploy machine learning models fast.

Amazon SageMaker is a fully managed service that provides every developer and data scientist with the ability to build, train, and deploy ML models at scale. It removes the complexity from each step of the ML workflow so you can more easily deploy your ML use cases, anything from predictive maintenance to computer vision to predicting customer behaviors.



1. Amazon EC2: a web service that provides resizable compute capacity in the cloud. It is designed to make web-scale computing easier for developers.

## Goals:

1. Learn how to create and run machine learning models on cloud.
2. Demonstrate how to do data analysis with Amazon SageMaker Studio and explore the functions of SageMaker.
3. I will use a dataset from Kaggle as a sample to test the performance of SageMaker Studio.

## Dataset:

Videogamesales:

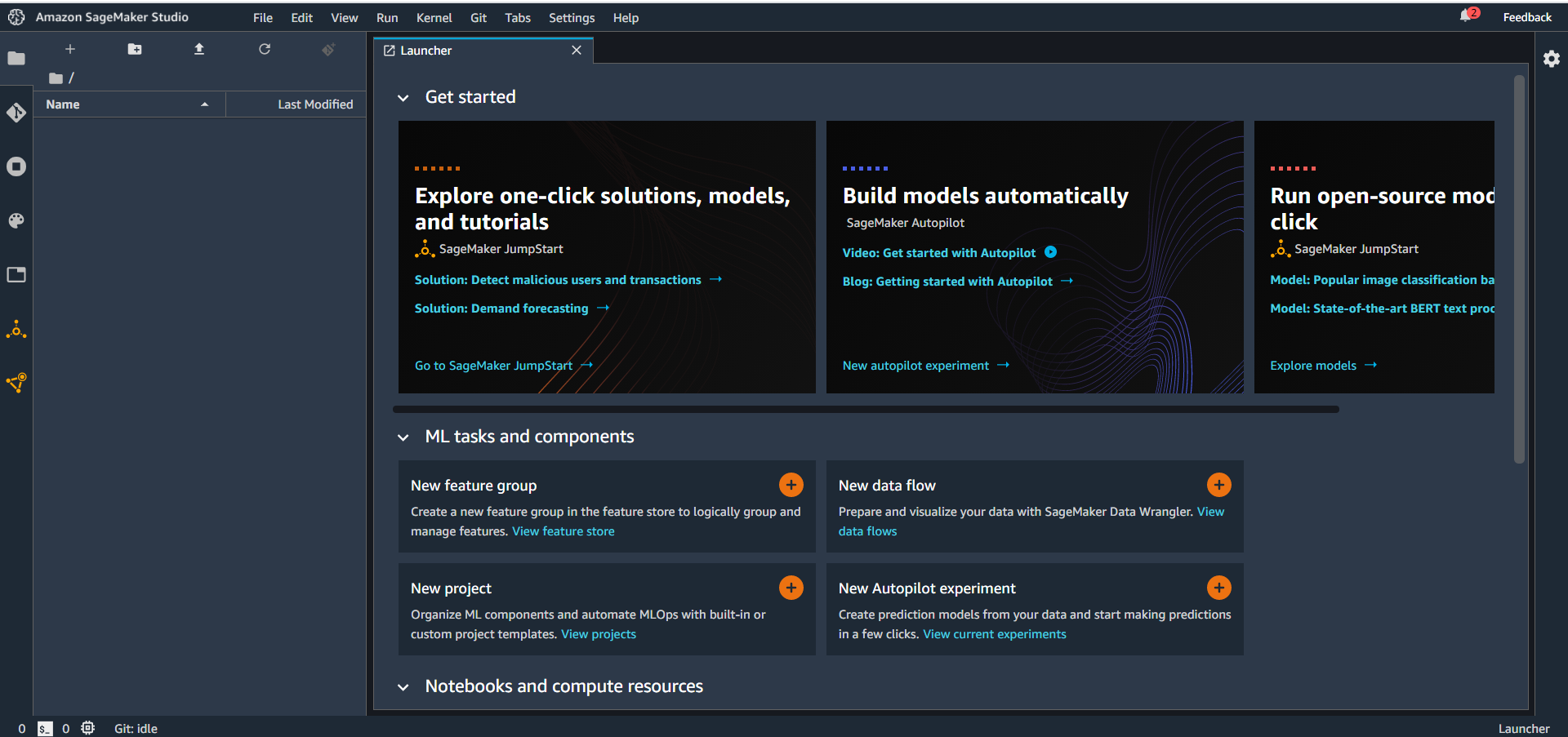
This dataset contains a list of video games with sales greater than 100,000 copies. It was generated by a scrape of vgchartz.com. It includes 16,598 records with 11 attributes.

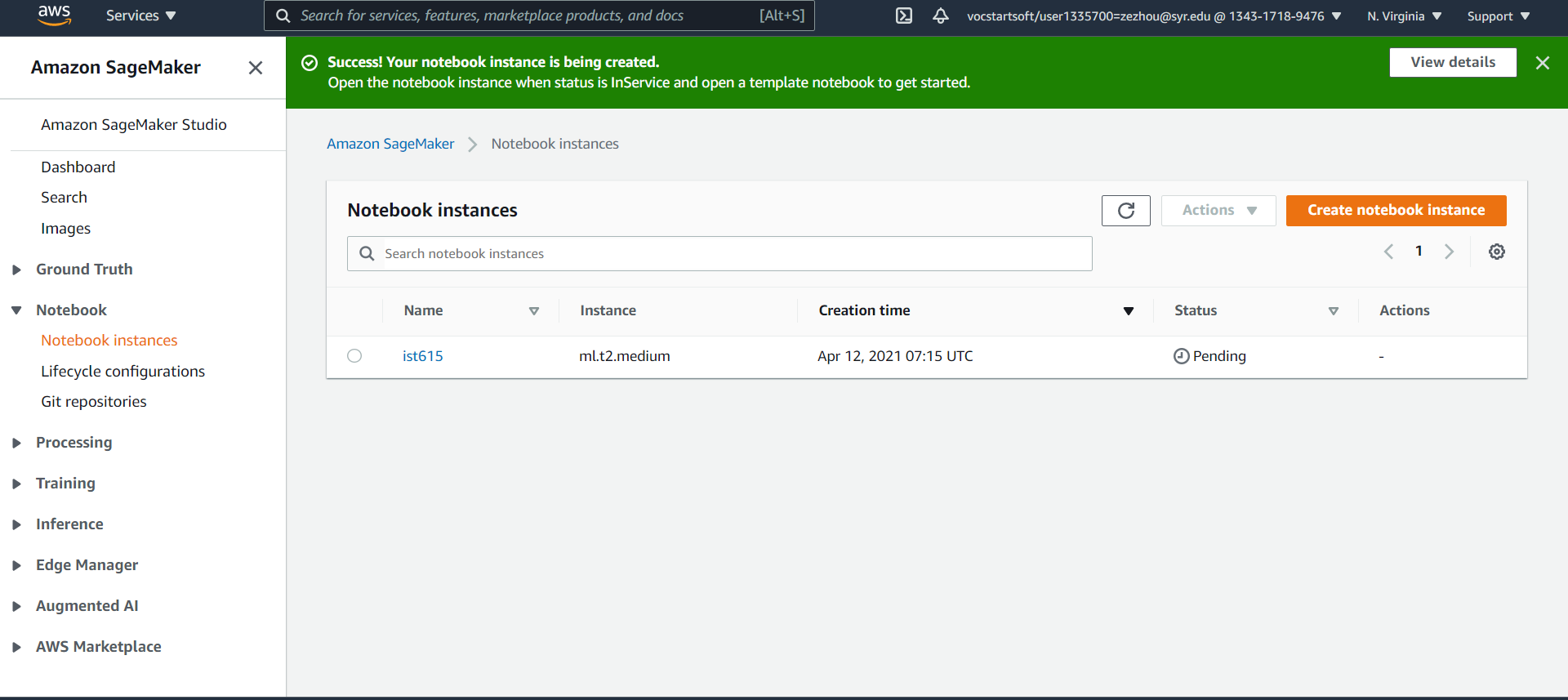


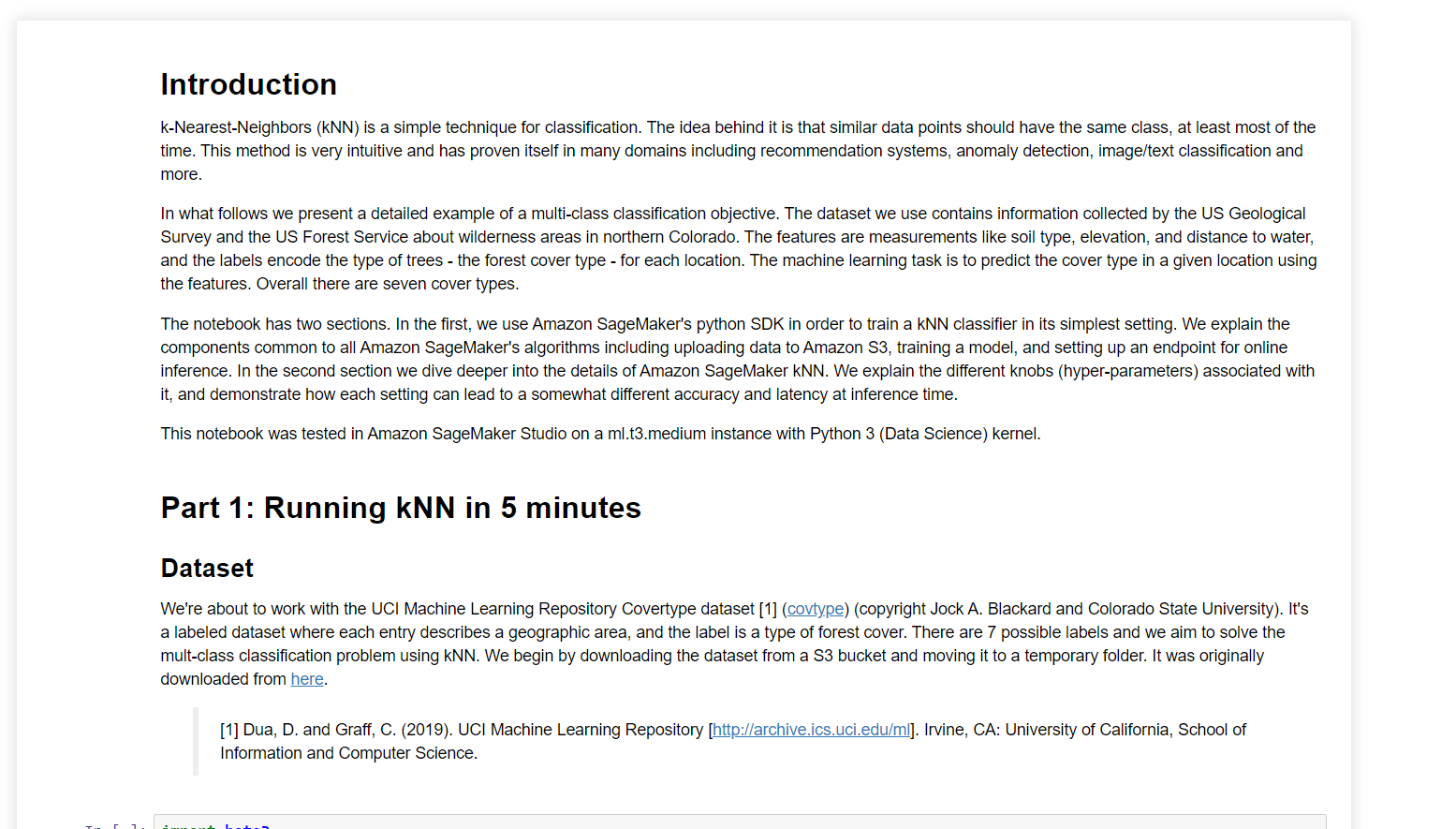
Link: <https://www.kaggle.com/gregorut/videogamesales>

## Reference:

(Explore AWS AI and Machine Learning Services, n.d.). Retrieved from: <https://aws.amazon.com/machine-learning/>







Questions:

What should I focus on? Machine learning (e.g., several different machine learning models, deep learning) or sagemaker studio (e.g., autopoint, autopilot, autoexperienment, a complete data analysis process of the dataset?

What’s important in my final report.