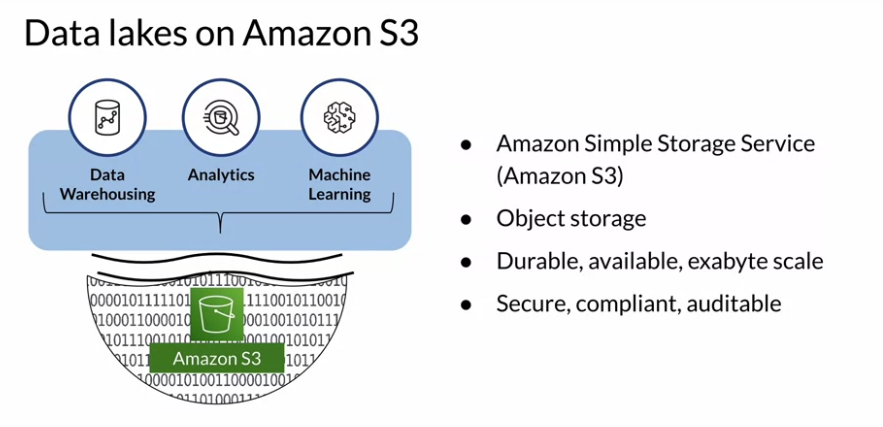
## [Analyze Datasets and Train ML Models using AutoML](https://www.coursera.org/learn/automl-datasets-ml-models/home/welcome)

**Week 1 - Explore the Use Case and Analyze the Dataset**

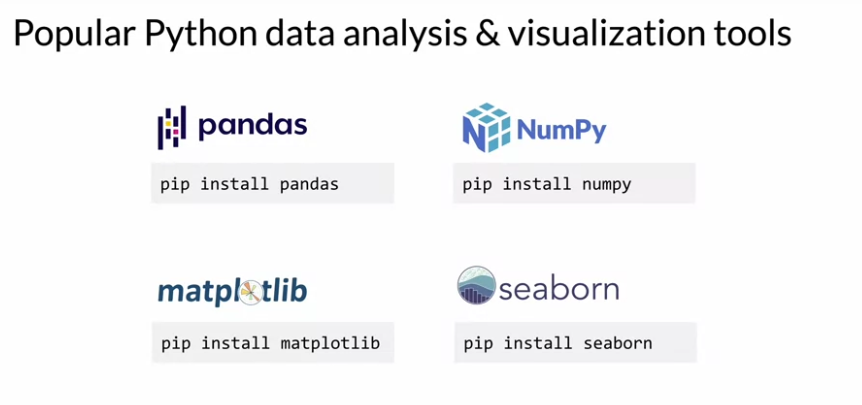
This week they talked about data ingestion, exploration and visualization.

**Data Ingestion & Exploration**



The data can be stored in S3, then with AWS Glue it will be made available for AWS Athena.

**Data Visualization**



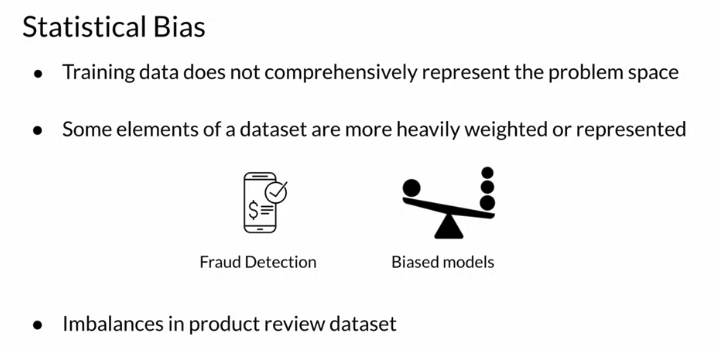
There are many ways and libraries to help with the visualization of your data.

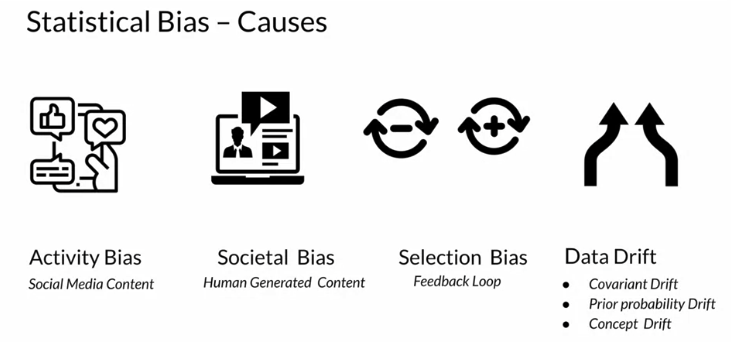
Example:

* For visualization: histograms, series plots, and so on.
* For libraries: pandas, numpy, matploblib and seaborn.

**Week 2 - Data Bias and Feature Importance**

This week they talked about Statistical Bias, Bias Detection and Feature Importance.





They elaborate two metrics for imbalance:

Class Imbalance (CI): measures the imbalance in the number of members between different facet values. E.g. "Does a product\_category have disproportionately more reviews than others?"

Difference in Proportions of Labels (DPL): measures the imbalance of positive outcomes between different facet values. E.g. "Does a product\_category have disproportionately higher ratings than others?"

For **detecting statistical bias** they present two approaches/tools: **Amazon SageMaker Data Wrangler** and **Amazon SageMaker Clarify**.

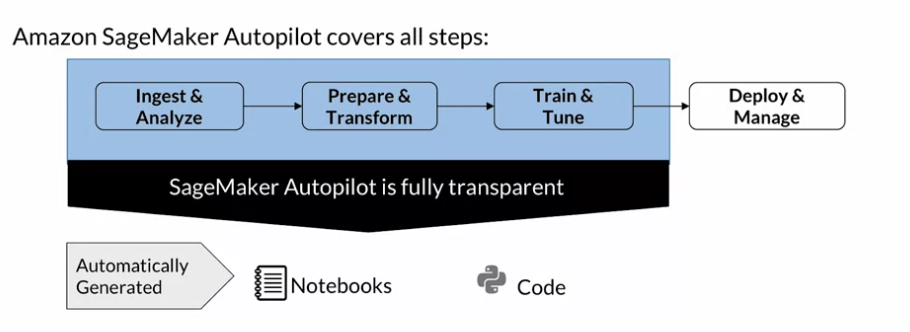
**Week 3 - Train a model with Amazon SageMaker Autopilot**

This week they talked about AutoML (the concept), what it comprehends and also about AutoPilot (the Automated Machine Learning tool from AWS).

AutoML will cover:

1. Ingest & Analyze: something like EDA and bias detection.
2. Prepare & Transform: data preparation.
3. Train & Tune: train different algorithms, evaluate them and compare.
4. Deploy & Manage: deploy the select model to an endpoint API.



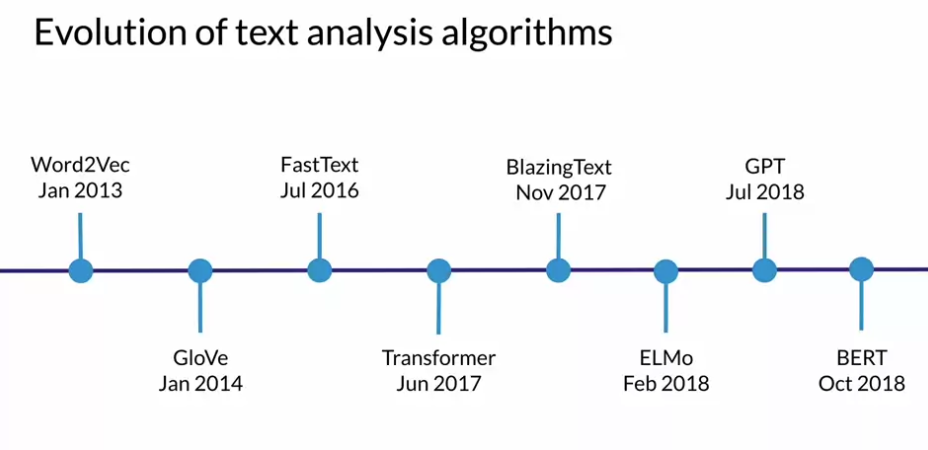


**Week 4 - Built-in algorithms**

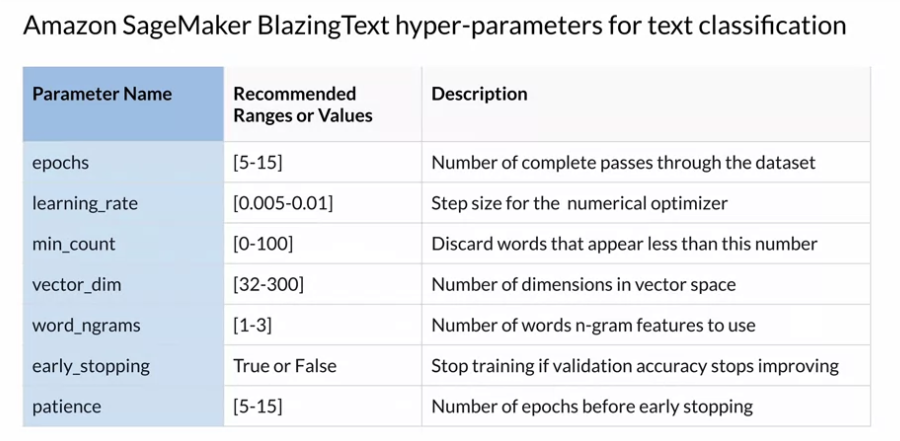
This week they talked about built-in algorithms, how they worked and different problem spaces:

* Classification & Regression (supervised)
* Clustering (unsupervised)
* Image Processing (computer vision)
* Text Analysis (NLP)

The focus was on Text Analysis algorithms with some brief explanation of each concept over time.



They also talked about Amazon SageMaker BlazingText and their parameters.



They also talked about how to deploy a model and provided snippets.

