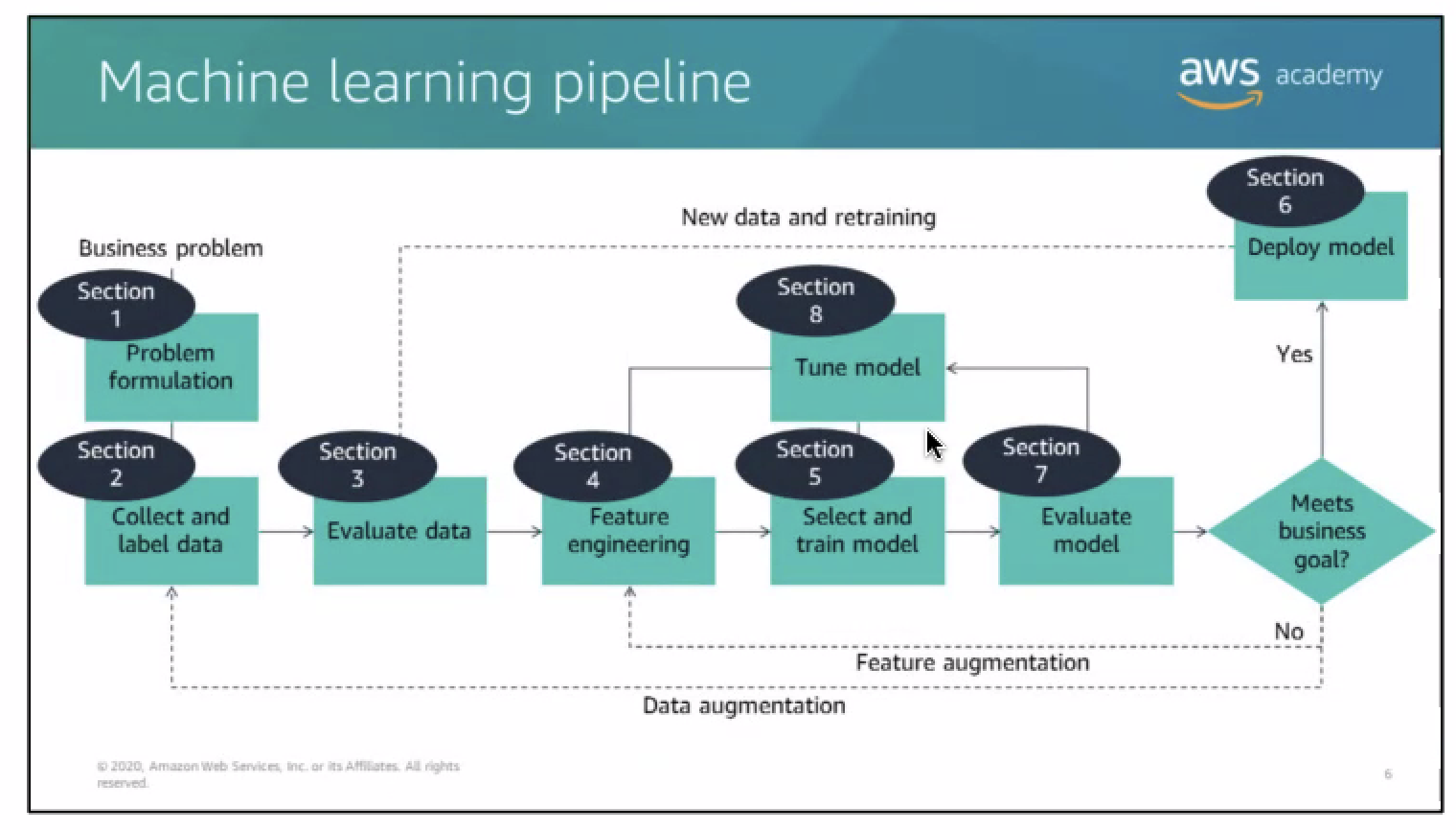
**IDS 706 Final Project Framework**

**Goal**

* Integrate and apply the analytic and modeling skills learned this semester



*[AWS 1143 module 2 & 3](https://awsacademy.instructure.com/courses/1143/modules" \l "module_14628)*

**Datasource**

* IPUMS, IMDBWS, Kyle’s Recommendation

**Skills Required**

* Data Infrastructure & Computation:
  + - * AWS 1098 (S3, Athena, Glue, Sagemaker)
      * AWS 1143 (ML Understanding )
* SQL: Datacamp, w3schools.com/sql/
* Data Manipulation: Pandas, Numpy
* Statistical Analysis & Visualization

**Milestones**

* Learning AWS 1143 (8th week)
* Data Preparation & Questions Formulation (9th week)
* Data Infrastructure (10th week)
* Data Modeling & Evaluation (11-12th week)
* Data Visualization (13th week)

**AWS Review**

1. S3: Upload .csv, encryption, query
2. Athena: Aggregate Multiple data sources into a database/table；借由query少部分的data/create partition来省钱; view 可用来hide complexity
3. Glue: Infer schema using crawler
4. Sagemaker: Use S3 cp command 将 s3 bucket的数据放到jupyter notebook来run analysis