

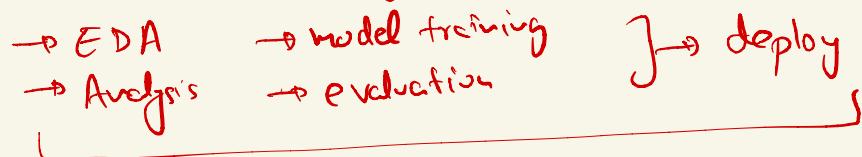

Aws Sagemaker

| Start at 9:10

→ Flask App → Docker → CI/CD → deploying it on ECS]

Streamlit → deploy it on streamlit Cloud]

→ Aws Sagemaker → Fully managed service



→ charge you a lot →

↳ model training

→ • deployment → expose the model as an endpoint

→ • GPUs, Big CPU, → Pre built models

◦ Automl, →

⇒ We need a place to store data →

Aws S3 → Storage

→ cheaper

- faster

• simple to use

) with other
aws services

Model

⇒ Banks → will a person buy fixed deposit not

↳ Xgboost library -

Notebook Type	Num CPUs	RAM	Price Per Month
ml.t2.medium	2	4	\$33.90
ml.t2.large	2	8	\$94.80
ml.t2.xlarge	8	32	\$189.70
ml.t3.medium	2	4	\$42.50
ml.t3.large	2	8	\$85.00
ml.t3.xlarge	4	16	\$170.10
ml.m4.xlarge	4	16	\$204.40
ml.m4.2xlarge	8	32	\$408.80
ml.m5.xlarge	4	16	\$196.40
ml.m5.2xlarge	8	32	\$392.70
ml.c4.xlarge	4	7.5	\$203.70
ml.c4.2xlarge	8	15	\$406.60
ml.p2.xlarge	4,1xK80 GPU	61	\$919.80
ml.p2.8xlarge	32, 8xK80 GPU	488	\$7358.40
ml.p3.2xlarge	8, 1xV100 GPU	61	\$207.00 Buffer

⇒ • How to Create S3 bucket from the code

- Specify which Region

- Modeling → xgboost

- deploying the model

↳ • How to access this model from anywhere
 ↳ localhost
 ↳ google colab

⇒ Similar to any normal machine

→ you can download data on SageMaker instance



Why the whole need of S3.

- To share with others
- Persistence

→ 16GB Ram 6/8 Core
 → 4 GB Ram 2 Core → 8 GB HDD
1TB 12TB

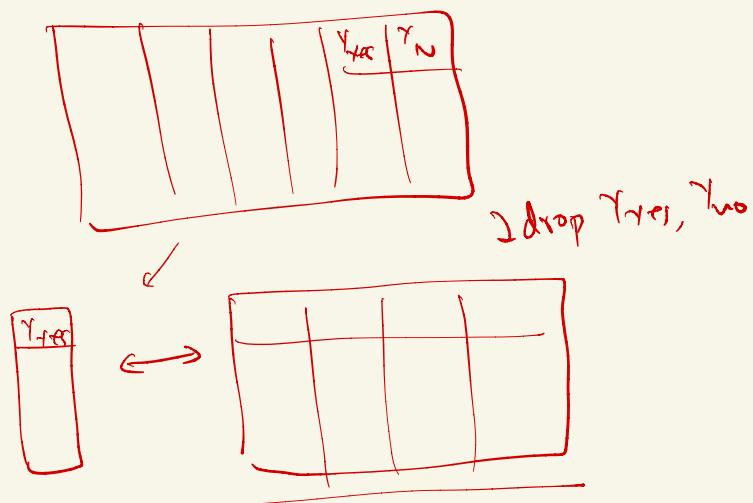
→ Segmeker expects the ^{training} data in a certain format →

Age	Mar	income	Gend	<u>Y</u>

Segmeker expects that the target variable is the first column of the data set

<u>Y</u>	Age	Gender		

⇒



- \Rightarrow
- Created a Sagemaker instance
 - ↳ Created an S3 bucket
 - Downloaded the data
 - ↳ Train
 - ↳ Test
 - ↳ Split
 - Moved the target variable ahead of all other columns
 - Sent/uploaded the training test data on S3
 - We have read data from S3

\Rightarrow popular \rightarrow painful to install

$\begin{matrix} \text{Prophet} \\ \text{xgboost} \end{matrix} \quad \left\{ \quad \rightarrow \quad \text{prebuilt which can be run} \right.$

Aws \rightarrow Store some model as ~~images~~ \rightarrow a app

\rightarrow Normal flow \rightarrow

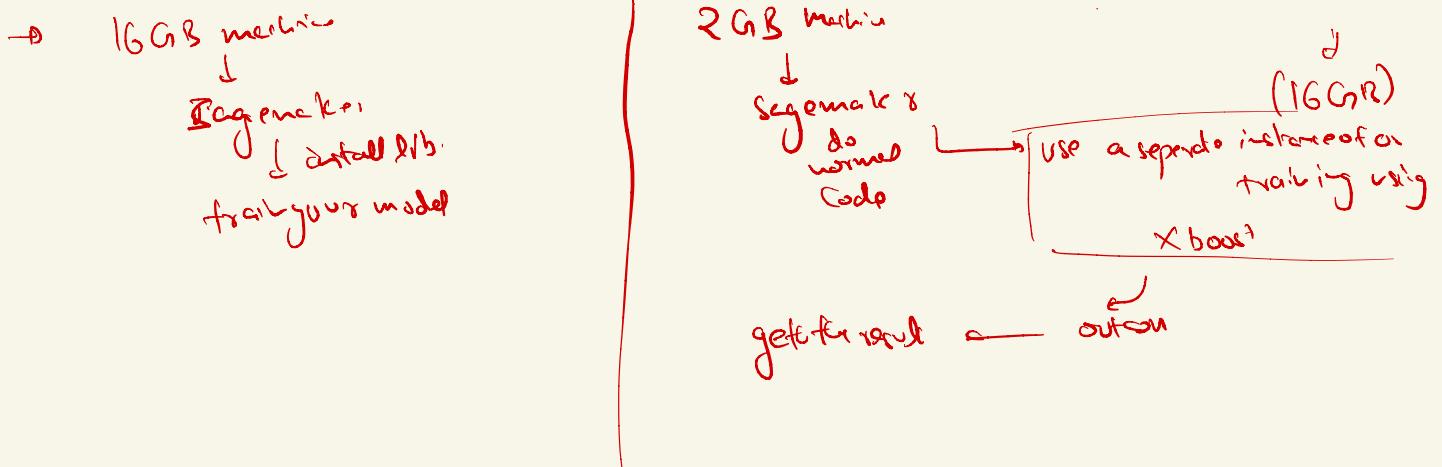
↳ install Libraries

- train your model

$\left| \begin{array}{l} \rightarrow \text{TensorFlow} \rightarrow \text{run popular model as a service} \\ \rightarrow \cdot \text{use an existing image} \rightarrow \text{give your data to this image} \\ \qquad \qquad \qquad \hookrightarrow \text{it will train the model.} \end{array} \right.$

\rightarrow XGBoost \rightarrow 16 GB RAM \rightarrow

\rightarrow Max memory goes at the time of model fitting/training.



→ → 1 instance → only require RAM for doing basic stuff

→ 1 instance → • available only during training
• Have higher RAM / GPU

→ 1 instance → • Instance where model is finally deployed
• Size of this instance depends on how much you want to use → this api

→ Cost is the major problem → Fully managed service
(→ You don't see all the things running background)

→ Cloudwatch → • resources
• billing

⇒ • Segmenter → • create notebook
↳ your models | Train
↳ deploy them

• Convert to model •

