

PROJECT CARD



PROJECT CARD

GOAL:

- *Build, train, test and deploy an AI/ML model to predict which employees are more likely to quit based on their features.*

TOOL:

- *AWS SageMaker Canvas*

PRACTICAL REAL-WORLD APPLICATION:

- *This project can be effectively used in Human Resources Departments to predict employee attrition and understand key factors that contribute to it.*

DATA:

• **INPUTS:**

- *Job Involvement, Education, Job Satisfaction, Performance Rating, Relationship Satisfaction, Work Life Balance*

• **OUTPUT:**

- *Attrition (binary), i.e.: employee stayed or left*



<https://publicdomainvectors.org/en/free-clipart/Employee-search-concept/83802.html>

<https://www.kaggle.com/pavansubhasht/ibm-hr-analytics-attrition-dataset>

INPUTS AND OUTPUTS

INPUTS

JOB INVOLVEMENT
EDUCATION
JOB SATISFACTION
PERFORMANCE RATING
RELATIONSHIP
SATISFACTION
WORK LIFE BALANCE



AI/ML MODEL



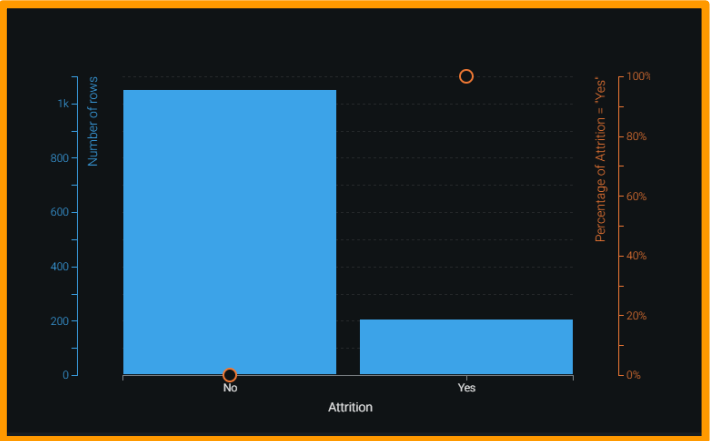
OUTPUT

ATTRITION
(0 or 1)

DATA OVERVIEW

Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber	EnvironmentSatisfaction	Gender	HourlyRate	JobInvolvement	JobLevel	JobRole	JobSatisfaction	MaritalStatus	MonthlyIncome	MonthlyRate	NumCompaniesWorked	Over18	OverTime	PercentSalaryHike	PerformanceRating
41	Yes	Travel_Rarely	1102	Sales	1	1	2 Life Sciences	1	1	2	Female	94	3	2	Sales Ex	4	Single	5393	13473	8	Y	Yes	11	3
4	No	Travel_Frequently	279	Research & Development	8	1	1 Life Sciences	1	2	4	Male	61	2	2	Research	2	Married	5130	24907	1	Y	No	23	4
31	Yes	Travel_Rarely	1373	Research & Development	2	2	2 Other	1	4	4	Male	92	2	1	Laborato	3	Single	2090	2396	6	Y	Yes	15	3
3	No	Travel_Frequently	1392	Research & Development	3	4	4 Life Sciences	1	5	4	Female	56	3	1	Research	3	Married	2309	23159	1	Y	Yes	11	3
21	No	Travel_Rarely	591	Research & Development	2	1	1 Medical	1	7	1	Male	40	3	1	Laborato	2	Married	3468	16632	9	Y	No	12	3
3	No	Travel_Frequently	1005	Research & Development	2	2	2 Life Sciences	1	8	4	Male	79	3	1	Laborato	4	Single	3068	11864	0	Y	No	13	3
53	No	Travel_Rarely	1324	Research & Development	3	3	3 Medical	1	10	3	Female	81	4	1	Laborato	1	Married	2670	9964	4	Y	Yes	20	4
3	No	Travel_Rarely	1358	Research & Development	24	1	1 Life Sciences	1	11	4	Male	67	3	1	Laborato	3	Divorced	2693	13335	1	Y	No	22	4
30	No	Travel_Frequently	216	Research & Development	23	3	3 Life Sciences	1	12	4	Male	44	2	3	Manufac	3	Single	9526	8787	0	Y	No	21	4
3	No	Travel_Rarely	1293	Research & Development	27	3	3 Medical	1	13	3	Male	94	3	2	Healthoa	3	Married	5237	16577	6	Y	No	13	3
3	No	Travel_Rarely	809	Research & Development	16	3	3 Medical	1	14	1	Male	84	4	1	Laborato	2	Married	2426	16479	0	Y	No	13	3
2	No	Travel_Rarely	153	Research & Development	15	2	2 Life Sciences	1	15	4	Female	49	2	2	Laborato	3	Single	4193	12682	0	Y	Yes	12	3
3	No	Travel_Rarely	670	Research & Development	26	1	1 Life Sciences	1	16	1	Male	31	3	1	Research	3	Divorced	2911	15170	1	Y	No	17	3
3	No	Travel_Rarely	1346	Research & Development	19	2	2 Medical	1	18	2	Male	93	3	1	Laborato	4	Divorced	2661	8758	0	Y	No	11	3
20	Yes	Travel_Rarely	103	Research & Development	24	3	3 Life Sciences	1	19	3	Male	50	2	1	Laborato	3	Single	2028	12347	5	Y	Yes	14	3
2	No	Travel_Rarely	1389	Research & Development	21	4	4 Life Sciences	1	20	2	Female	51	4	3	Manufac	1	Divorced	9980	10195	1	Y	No	11	3
30	No	Travel_Rarely	334	Research & Development	5	2	2 Life Sciences	1	21	1	Male	80	4	1	Research	2	Divorced	3298	15053	0	Y	Yes	12	3
2	No	Non-Travel	1123	Research & Development	16	2	2 Medical	1	22	4	Male	96	4	1	Laborato	4	Divorced	2935	7324	1	Y	Yes	13	3
5	No	Travel_Rarely	1219	Sales	2	4	4 Life Sciences	1	23	1	Female	78	2	4	Manager	2	Married	15427	22021	2	Y	No	16	3
3	No	Travel_Rarely	371	Research & Development	2	3	3 Life Sciences	1	24	4	Male	45	3	1	Research	4	Single	3944	4306	5	Y	Yes	11	3
2	No	Non-Travel	673	Research & Development	11	2	2 Other	1	26	1	Female	96	4	2	Manufac	3	Divorced	4011	8232	0	Y	No	18	3
3	Yes	Travel_Rarely	1218	Sales	9	4	4 Life Sciences	1	27	3	Male	82	2	1	Sales Re	1	Single	3407	6986	7	Y	No	23	4
3	No	Travel_Rarely	419	Research & Development	7	4	4 Life Sciences	1	28	1	Female	53	3	3	Research	2	Single	11994	21233	0	Y	No	11	3
2	No	Travel_Rarely	391	Research & Development	15	2	2 Life Sciences	1	30	3	Male	96	3	1	Research	4	Single	1232	19281	1	Y	No	14	3
3	Yes	Travel_Rarely	699	Research & Development	6	1	1 Medical	1	31	2	Male	83	3	1	Research	1	Single	2960	17102	2	Y	No	11	3
5	No	Travel_Rarely	1282	Research & Development	5	3	3 Other	1	32	3	Female	58	3	5	Manager	3	Divorced	19094	10735	4	Y	No	11	3
3	Yes	Travel_Frequently	1125	Research & Development	16	1	1 Life Sciences	1	33	2	Female	72	1	1	Research	1	Single	3919	4681	1	Y	Yes	22	4
4	No	Travel_Rarely	691	Sales	8	4	4 Marketing	1	35	3	Male	48	3	2	Sales Ex	2	Married	6825	21173	0	Y	No	11	3
4	No	Travel_Rarely	477	Research & Development	7	4	4 Medical	1	36	1	Female	42	2	3	Healthoa	4	Married	10248	2094	3	Y	No	14	3
4	No	Travel_Rarely	705	Sales	2	4	4 Marketing	1	38	2	Female	83	3	5	Manager	1	Single	18947	22622	3	Y	No	12	3
3	No	Travel_Rarely	924	Research & Development	2	3	3 Medical	1	39	4	Male	78	3	1	Laborato	4	Single	2496	6670	4	Y	No	11	3
4	No	Travel_Rarely	1459	Research & Development	10	4	4 Other	1	40	4	Male	41	3	2	Healthoa	4	Married	6465	19121	2	Y	Yes	13	3
3	No	Travel_Rarely	125	Research & Development	9	2	2 Medical	1	41	4	Male	83	2	1	Laborato	3	Single	2206	16117	1	Y	No	13	3
30	Yes	Travel_Rarely	895	Sales	5	3	3 Technical Degree	1	42	4	Male	56	3	2	Sales Re	4	Married	2086	3335	3	Y	No	14	3
2	Yes	Travel_Rarely	813	Research & Development	1	3	3 Medical	1	45	2	Male	61	3	1	Research	4	Married	2293	3020	2	Y	Yes	16	3
4	No	Travel_Rarely	1273	Research & Development	2	2	2 Medical	1	46	4	Female	72	4	1	Research	3	Divorced	2645	21923	1	Y	No	12	3
5	Yes	Travel_Rarely	869	Sales	3	2	2 Marketing	1	47	1	Male	86	2	1	Sales Re	3	Married	2683	3810	1	Y	Yes	14	3
35	No	Travel_Rarely	890	Sales	2	3	3 Marketing	1	49	4	Female	97	3	1	Sales Re	4	Married	2014	9687	1	Y	No	13	3
3	No	Travel_Rarely	852	Research & Development	5	4	4 Life Sciences	1	51	2	Female	82	2	1	Research	1	Married	3419	13072	9	Y	Yes	14	3
31	No	Travel_Frequently	1141	Sales	1	3	3 Life Sciences	1	52	3	Female	42	4	2	Sales Ex	1	Married	5376	3193	2	Y	No	19	3
3	No	Travel_Rarely	464	Research & Development	4	2	2 Other	1	53	3	Male	75	3	1	Laborato	4	Divorced	1951	10910	1	Y	No	12	3
27	No	Travel_Rarely	1240	Research & Development	2	4	4 Life Sciences	1	54	4	Female	33	3	1	Laborato	1	Divorced	2341	19715	1	Y	No	13	3
2	Yes	Travel_Rarely	1357	Research & Development	25	3	3 Life Sciences	1	55	1	Male	48	1	1	Laborato	3	Single	2293	10558	1	Y	No	12	3

MODEL OUTPUT: ATTRITION (BINARY VALUE) (YES OR NO)



PROJECT DEMO

?

≡

?

👤

>

Employee_attrition_prediction

V1 Ready Add version Share

Select

Build

Analyze

Predict

Predict target values

Batch prediction

Single prediction

Modify values to predict Attrition in real time.

Filter columns

Column	Feature importance ↓	Value
PerformanceRating	<div><div></div></div> 13.9%	<div>3</div>
MaritalStatus	<div><div></div></div> 9.8%	<div>Married</div>
NumCompaniesWorked	<div><div></div></div> 7.02%	<div>1</div>
RelationshipSatisfaction	<div><div></div></div> 5.49%	<div>3</div>
JobRole	<div><div></div></div> 4.71%	<div>Sales Executive</div>
DistanceFromHome	<div><div></div></div> 4.63%	<div>2</div>
JobLevel	<div><div></div></div> 4.62%	<div>1</div>
Age	<div><div></div></div> 2.14%	<div>35</div>

Attrition Prediction

Copy

No

Average prediction

No

55.8%

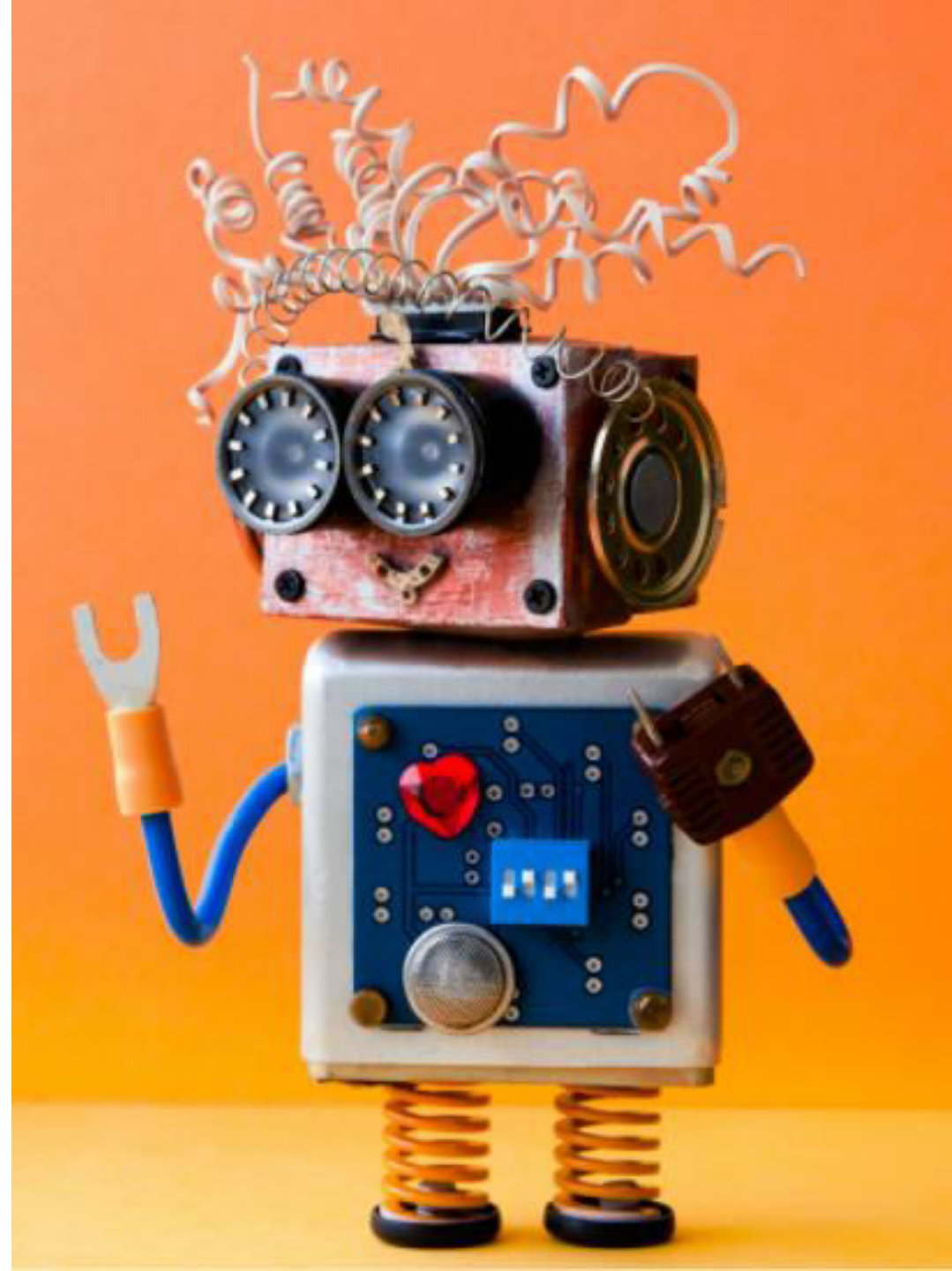
Yes

44.1%

Close

Download

WHAT IS SAGEMAKER CANVAS?



WHAT IS AWS SAGEMAKER CANVAS?

- AWS SageMaker Canvas empowers anyone to build, train and test a machine learning model without writing a single line of code!
 - With AWS Canvas, anyone can:
 - Import data from S3 or any other source
 - Build an AI/ML model
 - Assess model performance
 - Perform inference and generate predictions
 - Export model to SageMaker Studio
-
- AWS SageMaker Canvas Documentation:
<https://aws.amazon.com/sagemaker/canvas/>

SUCCESS STORIES



BUSINESS CASE AND SUCCESS STORIES

- Hiring/retaining employees is an expensive process that requires capital, time and skills.
- Average companies lose 1% - 2.5% of their total revenue on the time it takes to bring a new hire up to speed.
- Companies spend 15%-20% of the employee's salary to recruit a new candidate.
- 7 ways AI is reinventing Human Resources:
<https://www.cmswire.com/digital-workplace/7-ways-artificial-intelligence-is-reinventing-human-resources/>

7 Ways Artificial Intelligence Is Reinventing Human Resources



By Dom Nicastro | May 18, 2020



CHANNEL: Digital Workplace

BUSINESS CASE AND SUCCESS STORIES

**EXTRACT
INFORMATION FROM
CANDIDATES
RESUMES USING AI**

**ASSESS EMPLOYEE
REFERRALS**

**GAIN KEY DATA
INSIGHTS**

**IMPROVED
ENGAGEMENT USING
AI-POWERED
CHATBOTS**

**CUSTOMIZED
LEARNING JOURNEY
FOR EMPLOYEES**

**PREDICT
EMPLOYEES'
RETENTION,
POTENTIAL AND
FATIGUE**

**ENHANCED
WORKPLACE
ANALYTICS**

<https://www.cmswire.com/digital-workplace/7-ways-artificial-intelligence-is-reinventing-human-resources/>

READING TIME & QUIZ – HOW AI IS TRANSFORMING HUMAN RESOURCES?

- Please read the article below and answer the following quiz.
 - Link to Article: https://www.business-standard.com/article/jobs/employee-attribution-how-ai-is-transforming-human-resource-practices-118062701396_1.html

Employee attrition? How AI is transforming human resource practices

According to the IT major, the use of the Watson platform has also made talent acquisition much more accurate and seamless

Topics

Automation | Human Resources | Ibm

Bibhu Ranjan Mishra & Debasis Mohapatra
Last Updated at June 28, 2018 09:02 IST

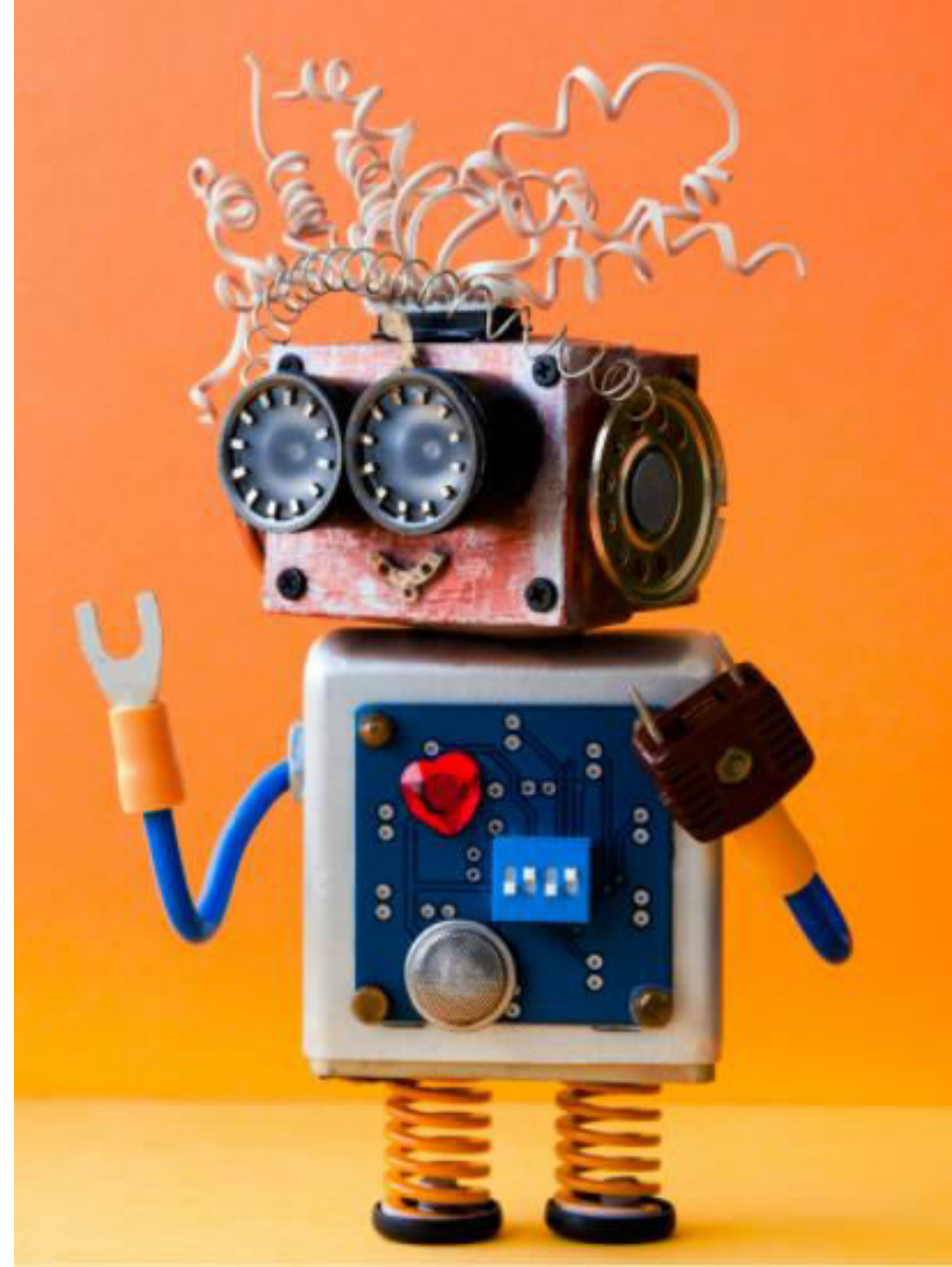


10 MINS



5 MINS

DEMO PART 1: LAUNCH CANVAS & UPLOAD DATA TO S3



AWS SAGEMAKER CANVAS SETUP

AFTER CREATING A SAGEMAKER DOMAIN, CLICK
ON LAUNCH APP AND THEN CANVAS

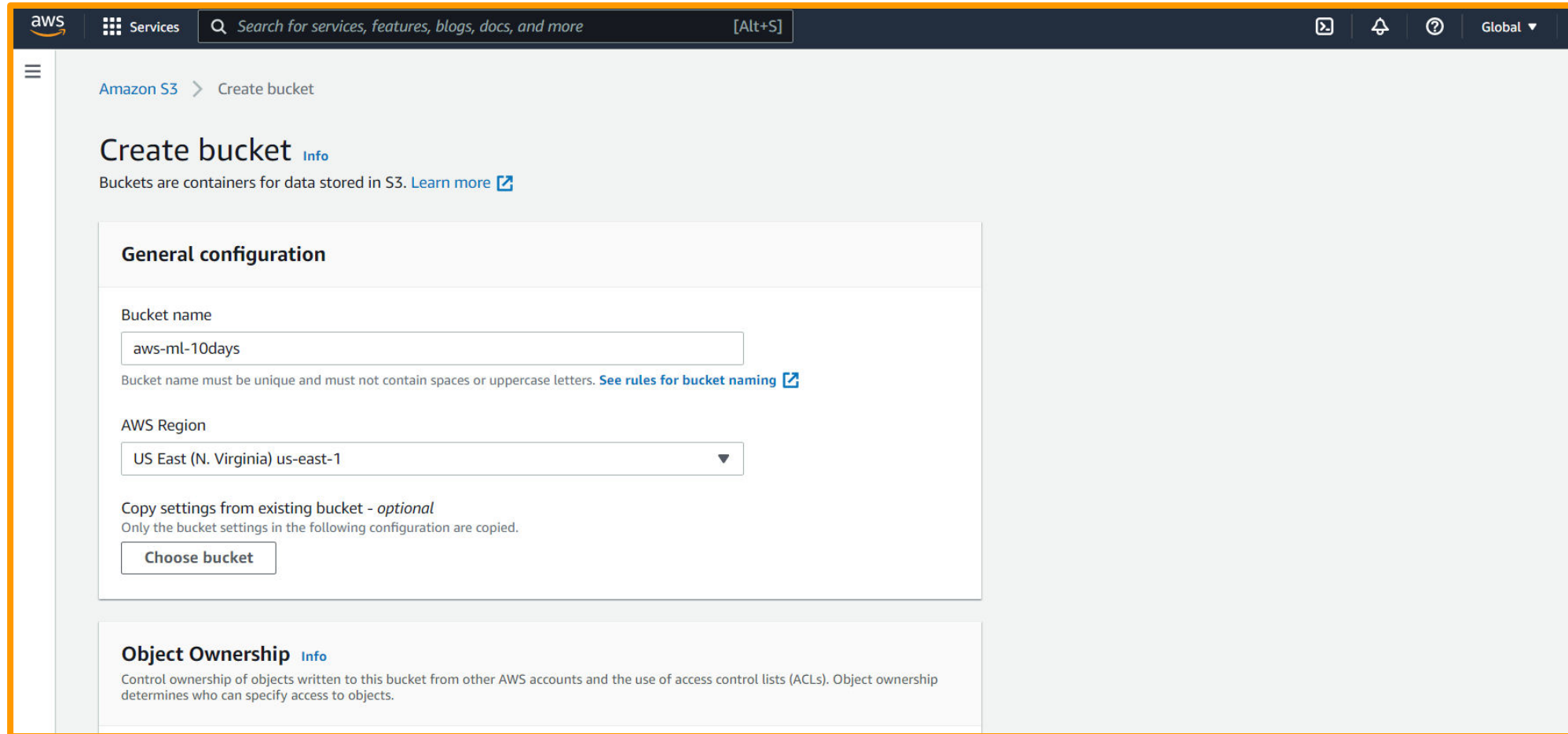
The screenshot displays the AWS SageMaker console interface. The top navigation bar includes the AWS logo, a 'Services' menu, a search bar, and regional/user information (N. Virginia, RyanAh). The left sidebar shows the 'Amazon SageMaker' header and a navigation menu with options: Dashboard, Search, SageMaker Domain (highlighted), Studio, RStudio, Canvas, Images, and a list of ML tasks (Ground Truth, Notebook, Processing, Training, Inference). The main content area is titled 'SageMaker Domain' and shows the configuration for a domain named 'default-1642779306981'. It includes a 'Users' section with an 'Add user' button and a search bar. Below this is a 'Domain' section with a 'Launch app' button and a dropdown menu showing 'Studio' and 'Canvas'. The 'Domain' section also contains a table with the following details:

▼ Domain	How to delete the domain	Delete Domain
Status ✔ Ready The status of the SageMaker Domain, and is not the status of the compute resources such as EC2 instances to execute notebook.	Domain ID d-z8peq3pv64pc Use the SageMaker Domain ID for troubleshooting and tracking usage.	Execution role arn:aws:iam::422132866096:role/service-role/AmazonSageMaker-ExecutionRole-20220121T103514
Authentication method AWS Identity and Access Management (IAM)		

At the bottom, a note states: 'Use Domain for troubleshooting and tracking usage. The status shown is for the SageMaker Studio service, and is not the status of compute resources such as EC2 instances to execute notebooks.'

AWS SAGEMAKER CANVAS SETUP

CREATE A NEW S3 BUCKET BEFORE YOU LAUNCH THE AWS
SAGEMAKER CANVAS



The screenshot shows the AWS Management Console interface for creating a new S3 bucket. The top navigation bar includes the AWS logo, a 'Services' menu, a search bar with the placeholder text 'Search for services, features, blogs, docs, and more', and a '[Alt+S]' keyboard shortcut. On the right side of the navigation bar are icons for a document, a bell, a question mark, and a 'Global' dropdown menu. The main content area has a breadcrumb trail 'Amazon S3 > Create bucket'. The title 'Create bucket' is followed by an 'Info' link. Below the title is a descriptive sentence: 'Buckets are containers for data stored in S3. [Learn more](#)'. The 'General configuration' section contains three main fields: 'Bucket name' with a text input containing 'aws-ml-10days', 'AWS Region' with a dropdown menu set to 'US East (N. Virginia) us-east-1', and a section for 'Copy settings from existing bucket - optional' which includes a 'Choose bucket' button. A note states 'Only the bucket settings in the following configuration are copied.' The 'Object Ownership' section is partially visible at the bottom, with an 'Info' link and a brief description of object ownership controls.

aws Services Search for services, features, blogs, docs, and more [Alt+S]

Amazon S3 > Create bucket

Create bucket [Info](#)

Buckets are containers for data stored in S3. [Learn more](#)

General configuration

Bucket name

Bucket name must be unique and must not contain spaces or uppercase letters. [See rules for bucket naming](#)

AWS Region

Copy settings from existing bucket - *optional*
Only the bucket settings in the following configuration are copied.

[Choose bucket](#)

Object Ownership [Info](#)

Control ownership of objects written to this bucket from other AWS accounts and the use of access control lists (ACLs). Object ownership determines who can specify access to objects.

AWS SAGEMAKER CANVAS SETUP

UPLOAD THE DATA TO THE S3 BUCKET

The screenshot displays the AWS S3 console interface. On the left, the 'Amazon S3' sidebar is visible, containing links to Buckets, Access Points, Object Lambda Access Points, Multi-Region Access Points, Batch Operations, and Access analyzer for S3. Below these, there's a section for 'Block Public Access settings for this account', followed by 'Storage Lens' with links to Dashboards and AWS Organizations settings. At the bottom of the sidebar is 'Feature spotlight' with a notification badge showing '3', and 'AWS Marketplace for S3'.

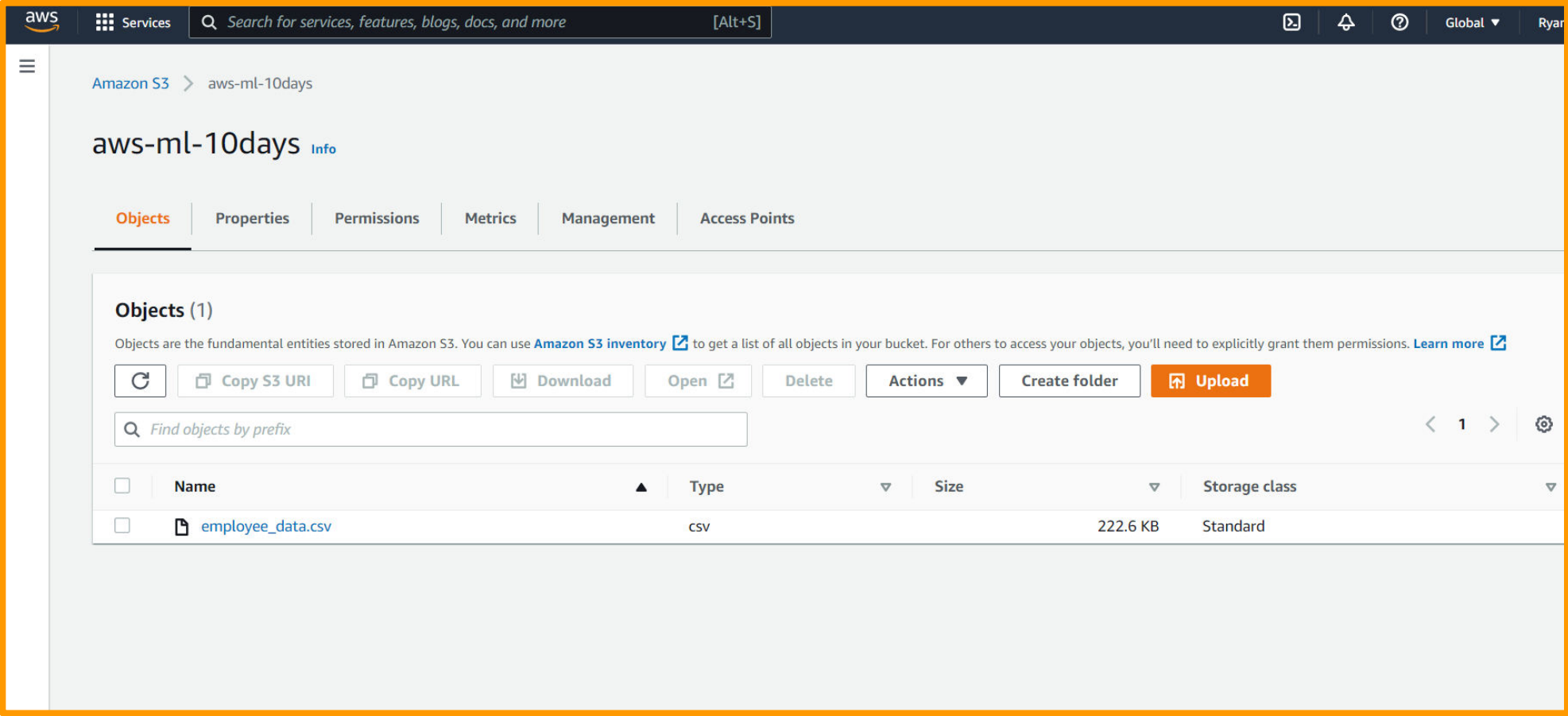
The main content area shows the 'aws-ml-10days' bucket. The breadcrumb navigation indicates 'Amazon S3 > aws-ml-10days'. The bucket name 'aws-ml-10days' is prominently displayed with an 'Info' link. Below the bucket name, there are tabs for 'Objects' (selected), 'Properties', 'Permissions', 'Metrics', 'Management', and 'Access Points'.

The 'Objects (0)' section contains a descriptive paragraph: 'Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)'. Below this text is a row of action buttons: a refresh icon, 'Copy S3 URI', 'Copy URL', 'Download', 'Open', 'Delete', 'Actions' (with a dropdown arrow), 'Create folder', and a highlighted 'Upload' button.

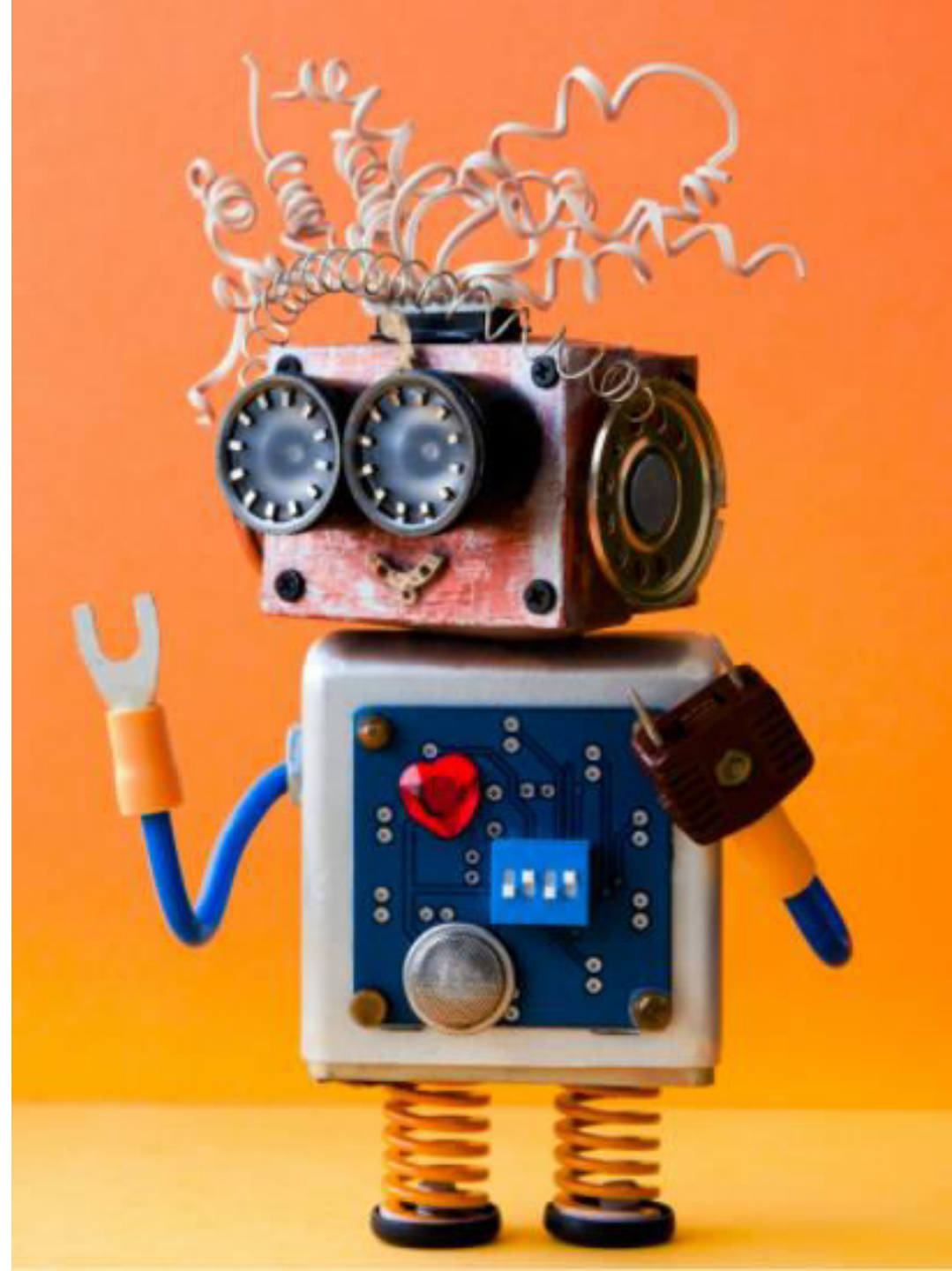
Below the buttons is a search bar with the placeholder text 'Find objects by prefix'. To the right of the search bar are navigation controls showing '< 1 >'. Below the search bar is a table header with columns: 'Name', 'Type', 'Last modified', 'Size', and 'Storage class'. The table body is empty, displaying the message 'No objects' and 'You don't have any objects in this bucket.' At the bottom of the table area is an 'Upload' button.

AWS SAGEMAKER CANVAS SETUP

EMPLOYEE_DATA.CSV IS NOW UPLOADED TO THE S3 BUCKET

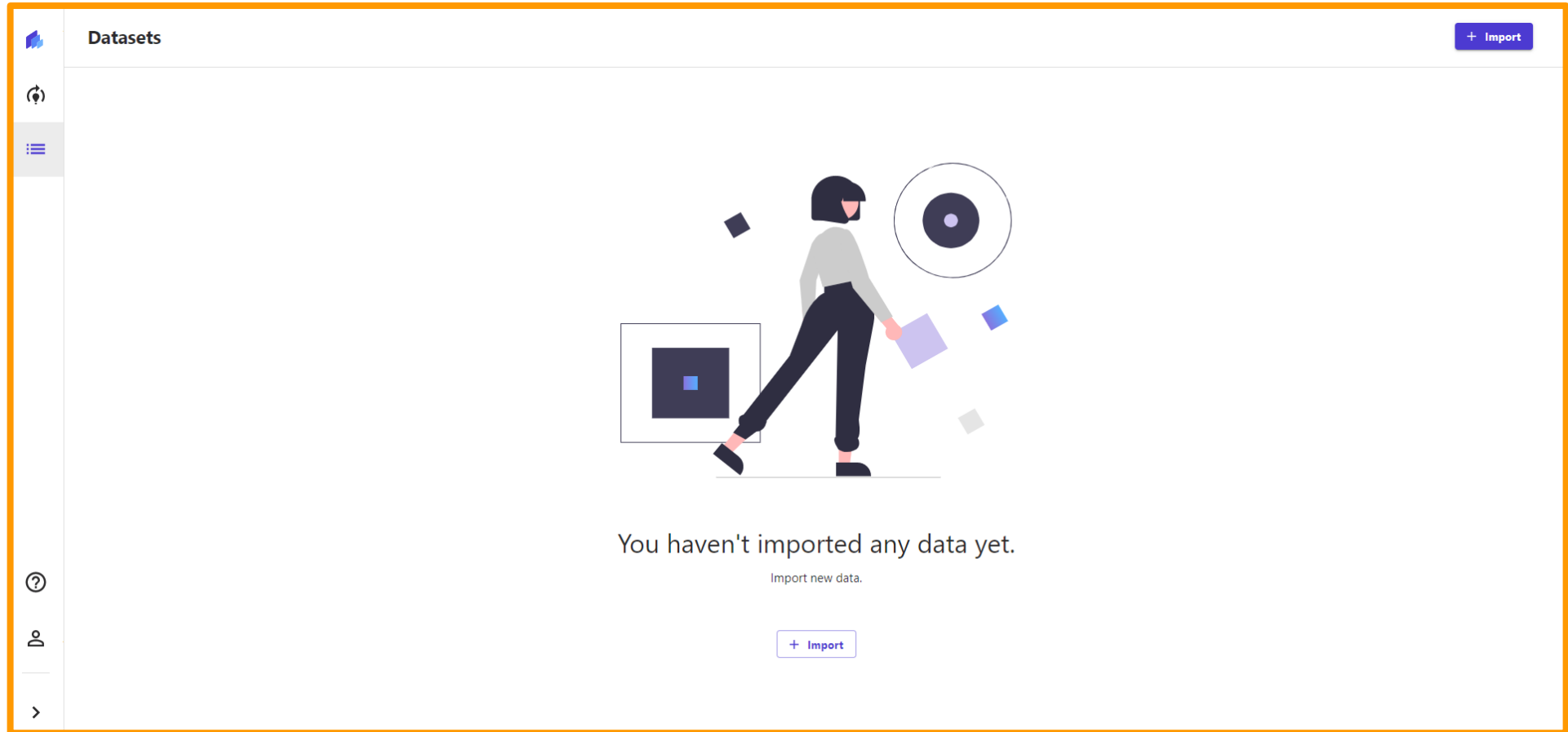


DEMO PART 2: DATA IMPORT INTO CANVAS & MODEL TRAINING



AWS SAGEMAKER CANVAS

FIRST YOU NEED TO IMPORT THE DATASET
FROM S3, CLICK ON IMPORT



AWS SAGEMAKER CANVAS

SELECT THE BUCKET FROM S3. VIEW AND
IMPORT THE DATASET

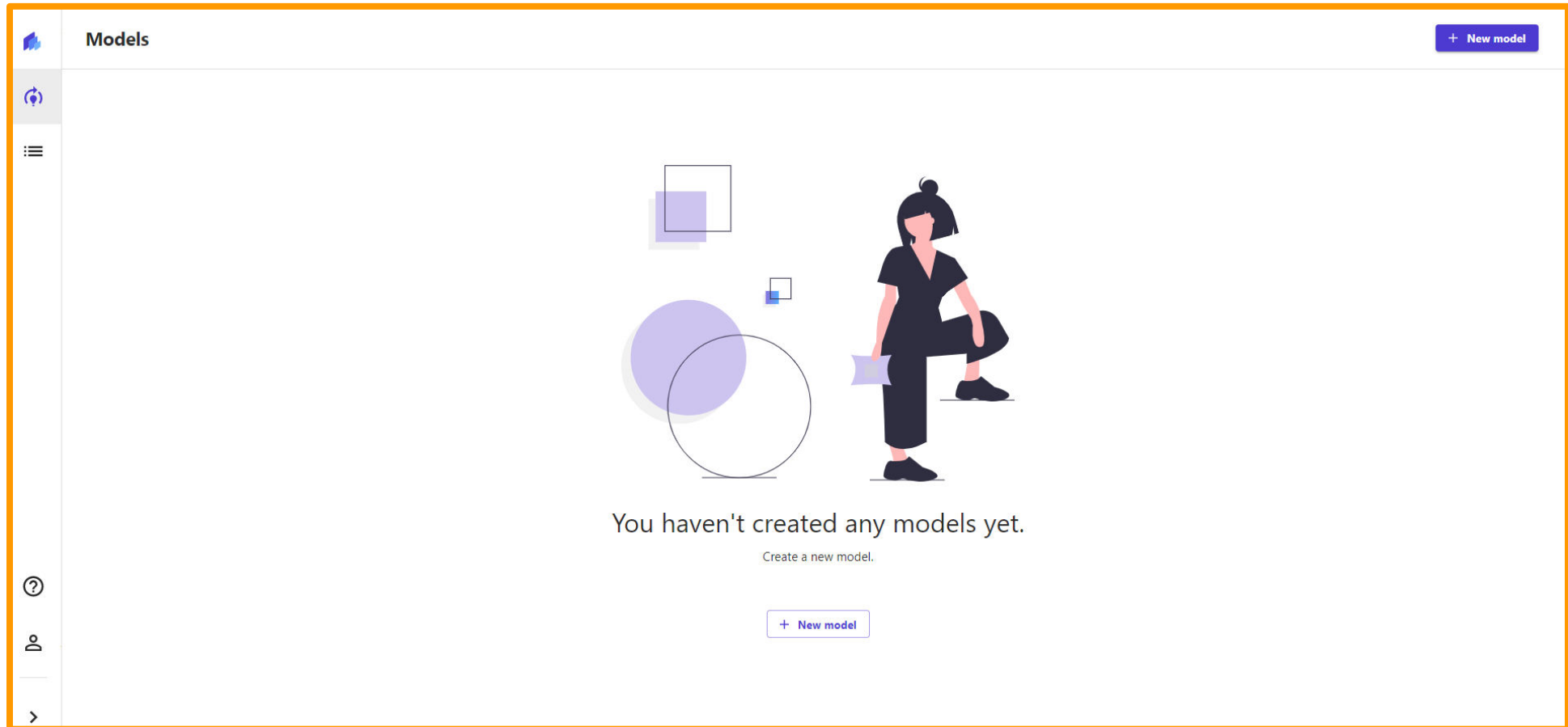
employee_data.csv Previewing first 100 rows

Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHo...	Education	EducationField	EmployeeCount	EmployeeNu...
41	Yes	Travel_Rarely	1102	Sales	1	2	Life Sciences	1	1
49	No	Travel_Frequently	279	Research & Developme...	8	1	Life Sciences	1	2
37	Yes	Travel_Rarely	1373	Research & Developme...	2	2	Other	1	4
33	No	Travel_Frequently	1392	Research & Developme...	3	4	Life Sciences	1	5
27	No	Travel_Rarely	591	Research & Developme...	2	1	Medical	1	7
32	No	Travel_Frequently	1005	Research & Developme...	2	2	Life Sciences	1	8
59	No	Travel_Rarely	1324	Research & Developme...	3	3	Medical	1	10
30	No	Travel_Rarely	1358	Research & Developme...	24	1	Life Sciences	1	11
38	No	Travel_Frequently	216	Research & Developme...	23	3	Life Sciences	1	12
36	No	Travel_Rarely	1299	Research & Developme...	27	3	Medical	1	13
35	No	Travel_Rarely	809	Research & Developme...	16	3	Medical	1	14
29	No	Travel_Rarely	153	Research & Developme...	15	2	Life Sciences	1	15
31	No	Travel_Rarely	670	Research & Developme...	26	1	Life Sciences	1	16

Select dataset

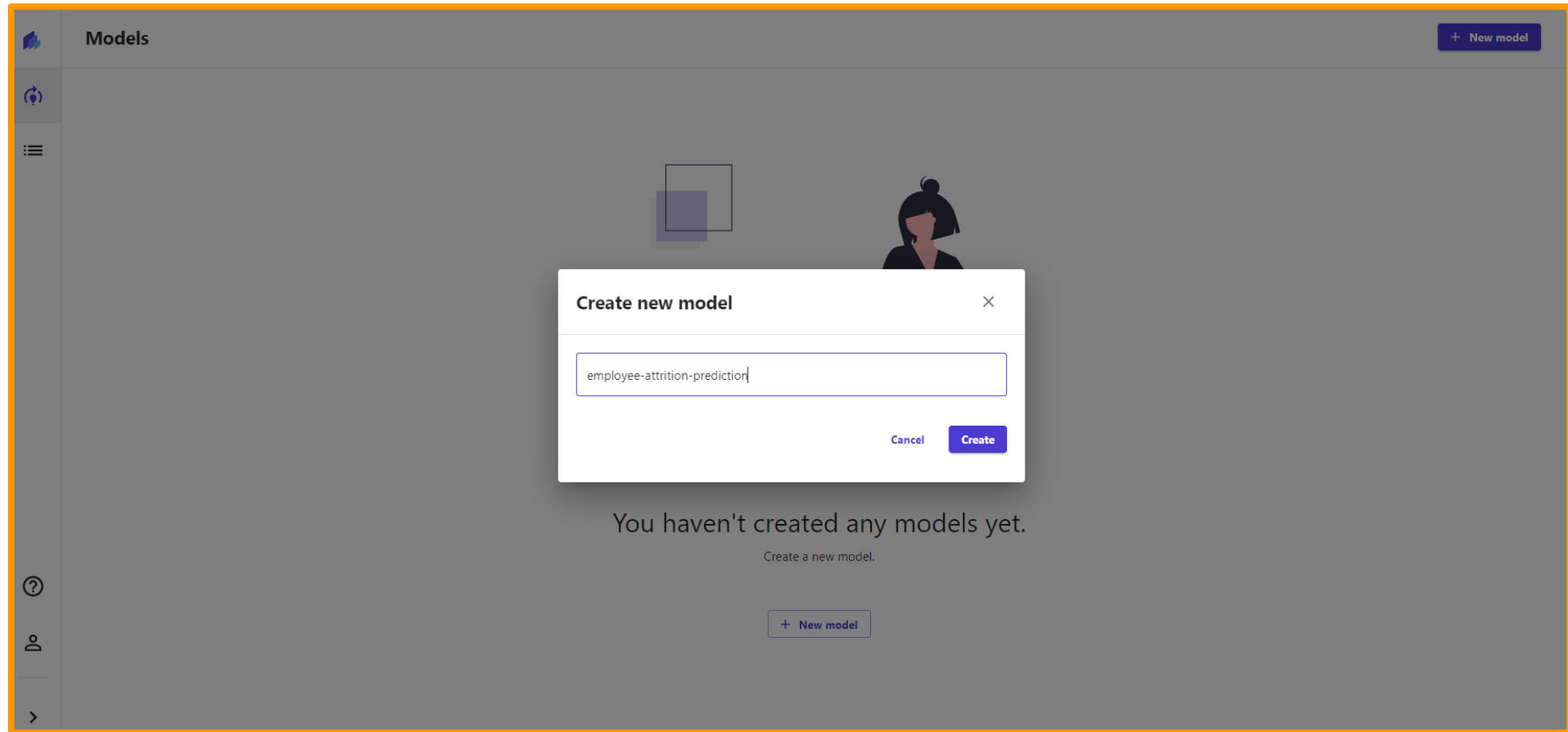
AWS SAGEMAKER CANVAS

CLICK ON MODELS AND CREATE A NEW
MODEL AND CLICK CREATE



AWS SAGEMAKER CANVAS

GIVE A NAME TO THE MODEL



AWS SAGEMAKER CANVAS

SELECT THE DATASET

employee-attribution-prediction

V1 • Draft

Add version

Share

Select

Build

Analyze

Predict

Select dataset

You can import a tabular dataset or choose one that has already been imported. Your dataset must contain at least one input column and a target column.

Search datasets in Canvas

AllJoined

Name	Source	Columns	Rows	Created	Status
<input checked="" type="radio"/> employee_data(2).csv	S3	35	1,470	01/21/2022 4:06 PM	Ready
<input type="radio"/> employee_data(2).csv	S3	35	1,470	01/21/2022 4:06 PM	Ready

CloseSelect dataset

AWS SAGEMAKER CANVAS

SELECT THE TARGET COLUMN “ATTRITION”,
CLICK ON “QUICK MODEL”

The screenshot shows the AWS SageMaker Canvas interface for a project named "employee-attribution-prediction". The interface is divided into four tabs: "Select", "Build", "Analyze", and "Predict". The "Build" tab is currently active.

Select a column to predict: A dropdown menu shows "Attrition" as the selected target column. Below it, a "Value distribution" bar chart shows two categories: "No" (blue bar) and "Yes" (orange bar).

Model type: The interface recommends "2 category prediction" for the selected target column. It states, "Your model classifies Attrition into two categories." A "Change type" link is available.

Quick build and Preview model: Two buttons are visible on the right side of the "Build" tab: "Quick build" (blue) and "Preview model" (orange).

Table of columns: Below the "Build" tab, a table lists the columns available for the dataset "employee_data(2).csv". The table has columns for "Column name", "Data type", "Missing", "Mismatched", "Unique", "Mean / Mode", and "Correlation to target".

<input type="checkbox"/>	Column name ↓	Data type	Missing ⓘ	Mismatched ⓘ	Unique ⓘ	Mean / Mode	Correlation to target ⓘ
<input checked="" type="checkbox"/>	YearsWithCurrManager	Numeric	0.00% (0)	0.00% (0)	18	2	-0.16
<input checked="" type="checkbox"/>	YearsSinceLastPromotion	Numeric	0.00% (0)	0.00% (0)	16	0	-0.03
<input checked="" type="checkbox"/>	YearsInCurrentRole	Numeric	0.00% (0)	0.00% (0)	19	2	-0.16
<input checked="" type="checkbox"/>	YearsAtCompany	Numeric	0.00% (0)	0.00% (0)	37	5	-0.13
<input checked="" type="checkbox"/>	WorkLifeBalance	Numeric	0.00% (0)	0.00% (0)	4	3	-0.06
<input checked="" type="checkbox"/>	TrainingTimesLastYear	Numeric	0.00% (0)	0.00% (0)	7	2	-0.06

At the bottom of the table, it shows "Total columns: 35" and "Total rows: 1,470". A "Close" button is located in the bottom right corner.

AWS SAGEMAKER CANVAS

QUICK MODEL TRAINING SHOULD TAKE 2-15
MINS TO TRAIN

employee-attribution-prediction V1 Building Add version Share

Select Build **Analyze** Predict

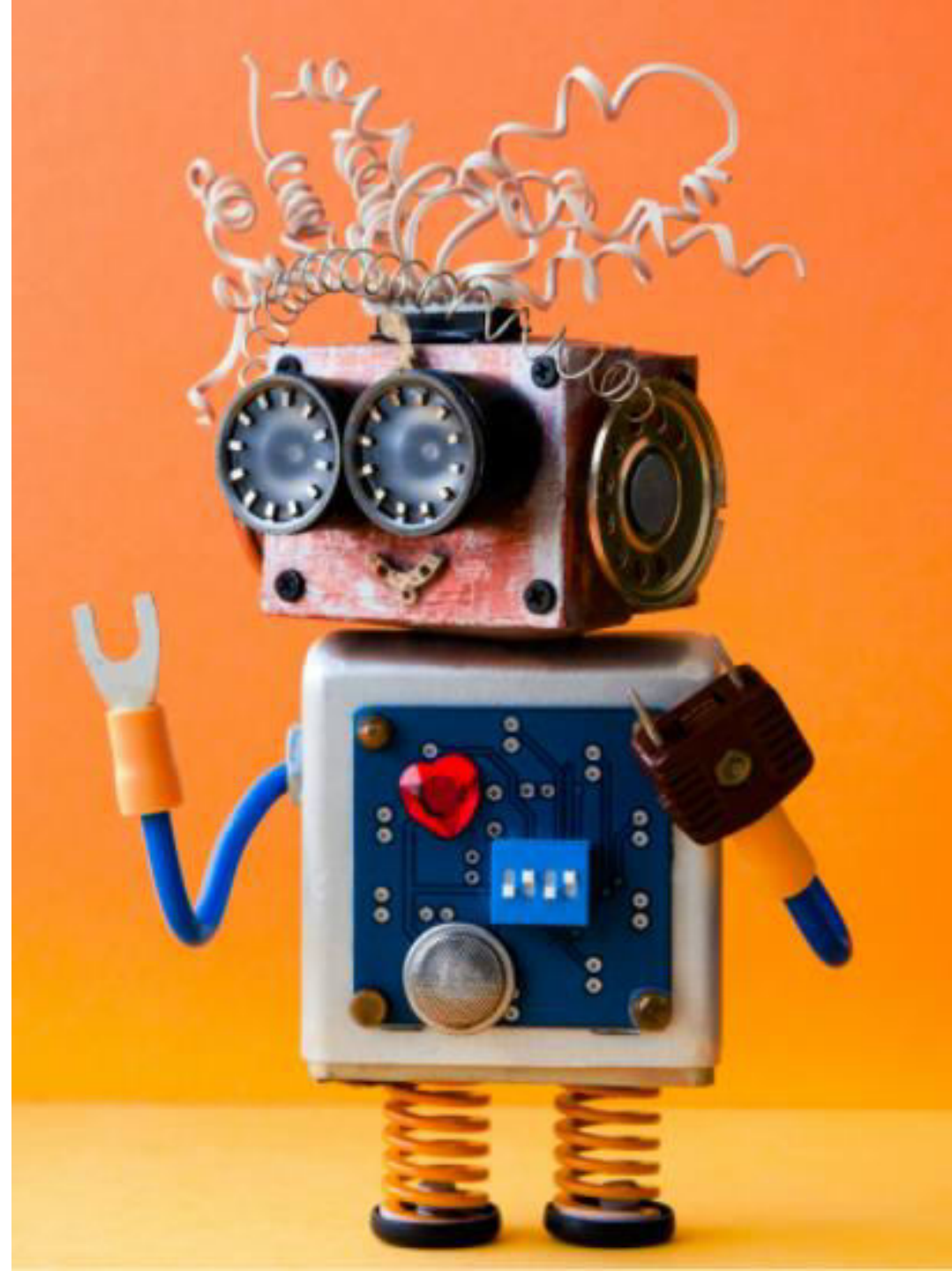
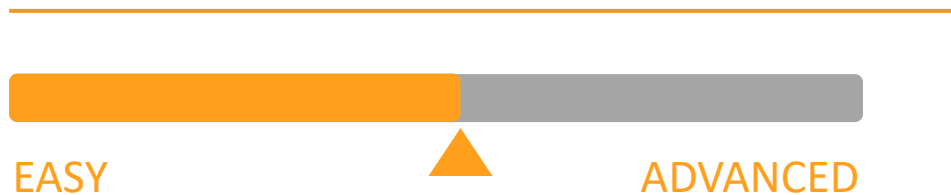
Model overview

Your model is being created. Quick build usually takes 2-15 minutes. You can now leave this view.

Expected build time	Build type	Detailed progress
2-15 minutes	Quick build	Generating column impact

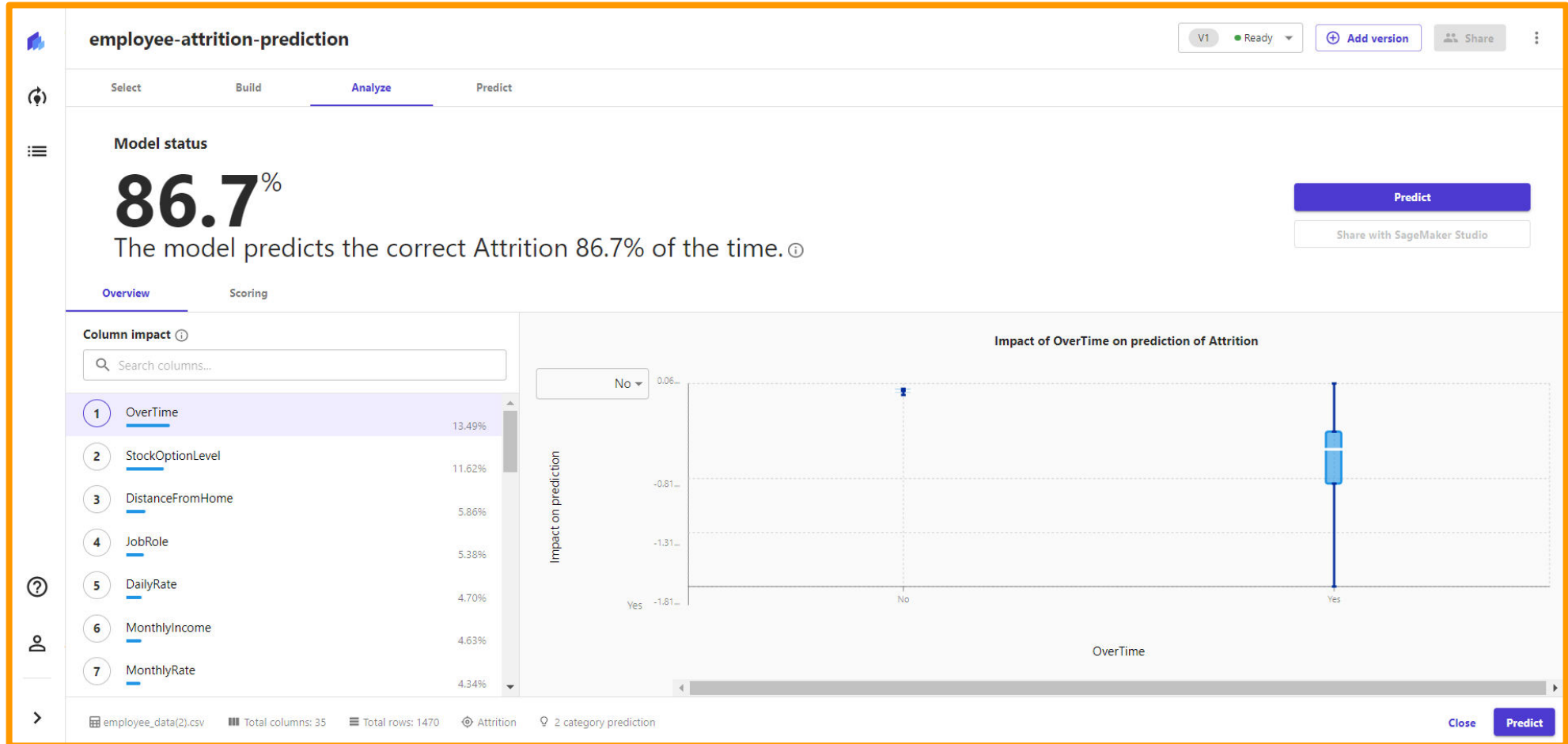
employee_data(2).csv Total columns: 35 Total rows: 1470 Attrition 2 category prediction Close

DEMO PART 3: TRAINED MODEL ASSESSMENT



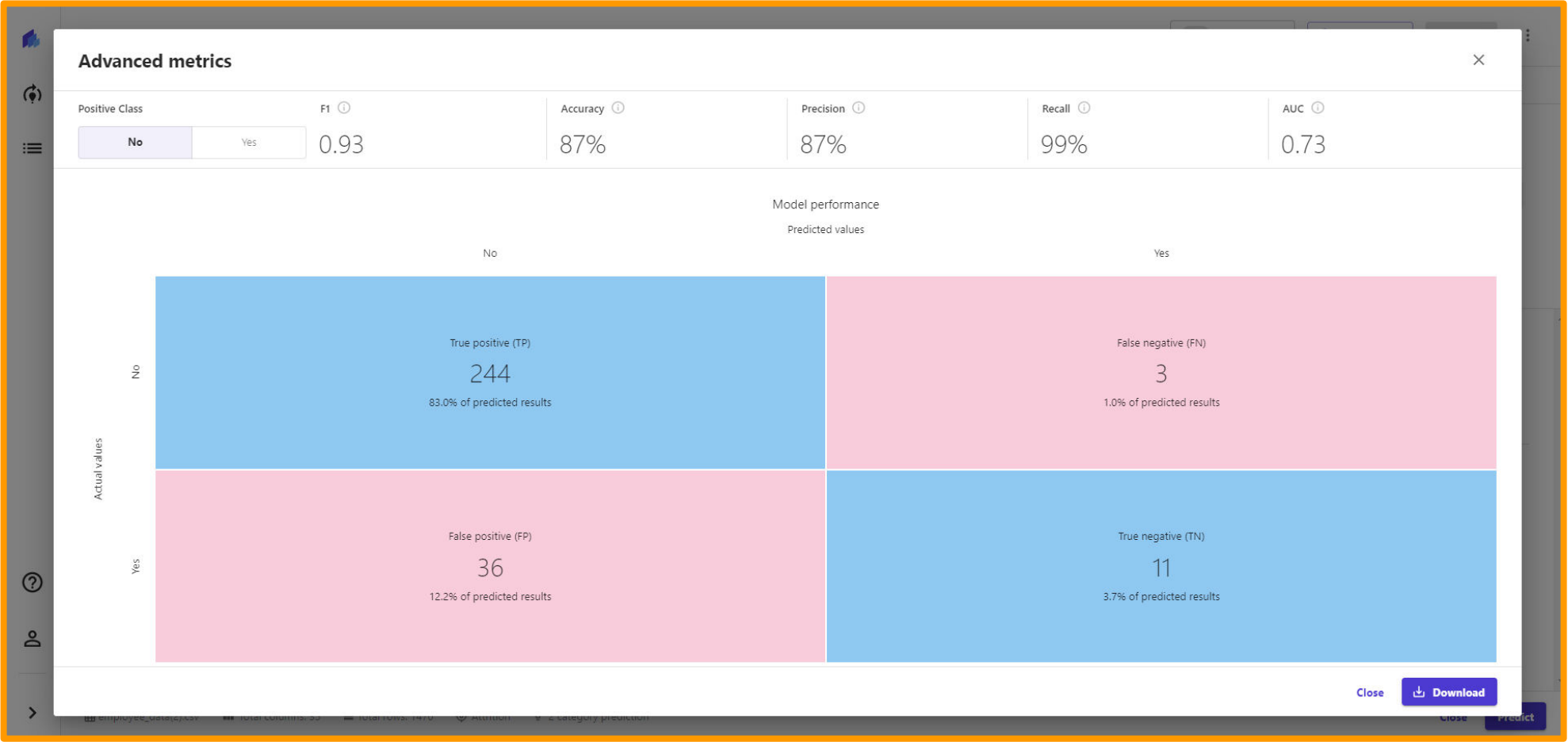
AWS SAGEMAKER CANVAS

AFTER THE MODEL IS TRAINED, IT'S NOW
READY TO BE ANALYZED! CHECK OUT THE
FEATURE IMPORTANCE (IMPACT)

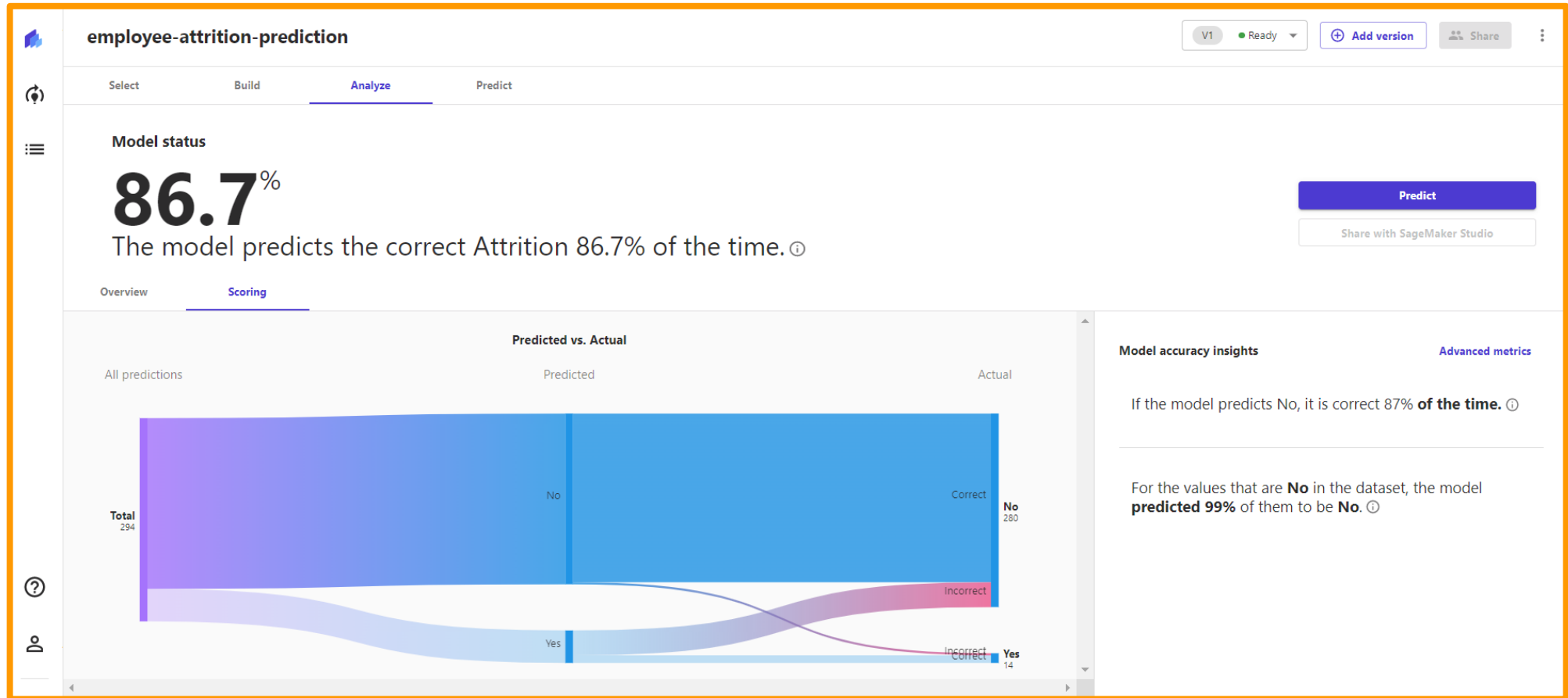


AWS SAGEMAKER CANVAS

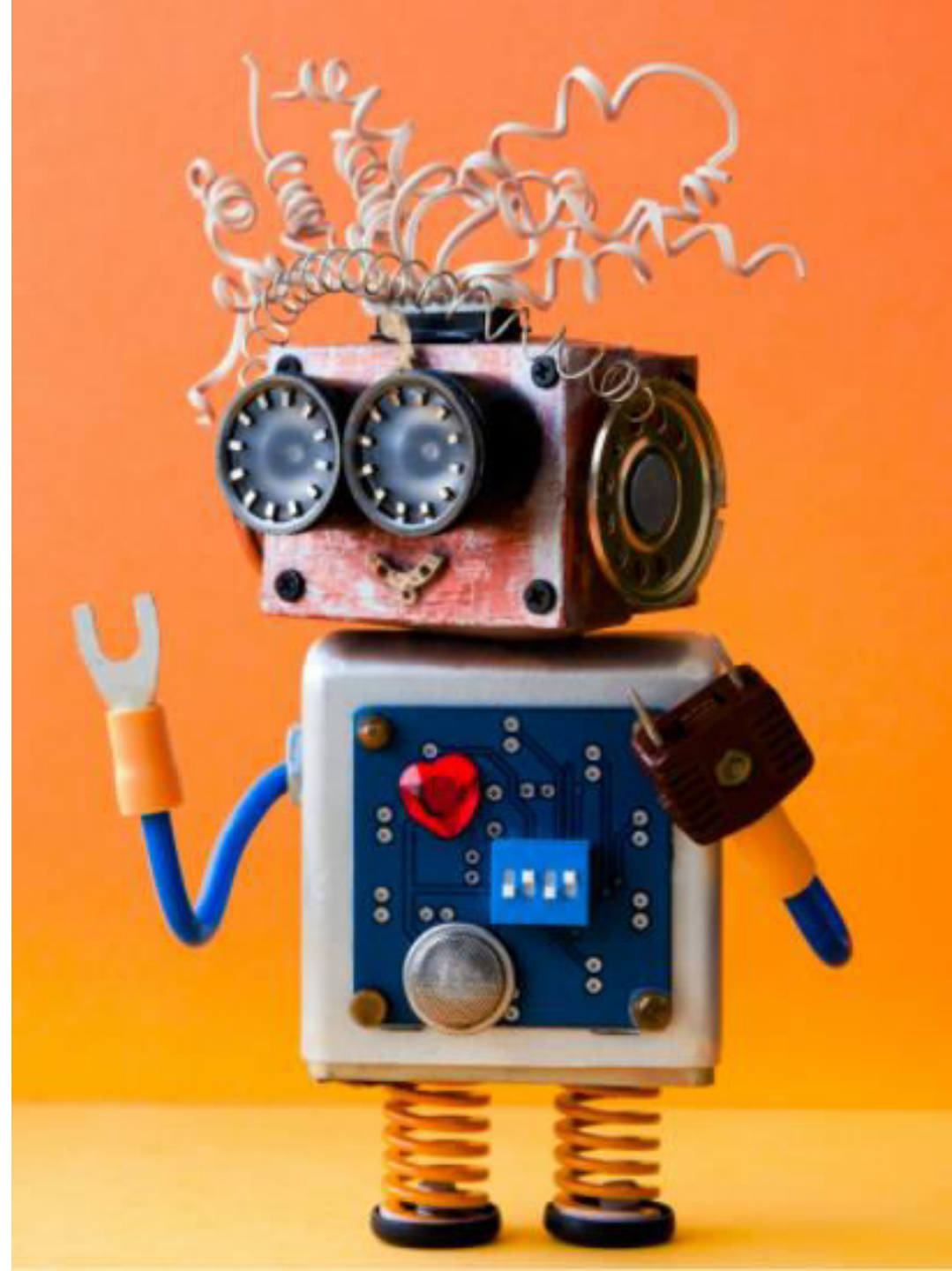
CONFUSION MATRIX



AWS SAGEMAKER CANVAS



DEMO PART 4: PERFORM MODEL INFERENCE



AWS SAGEMAKER CANVAS

SELECT INPUTS AND GENERATE OUTPUT

employee-attribution-prediction

V1 Ready + Add version Share

Select

Build

Analyze

Predict

Predict target values

Batch prediction

Single prediction

Modify values to predict Attrition in real time.

Filter columns

Column	Feature importance ↓	Value
MaritalStatus	<div></div> 16.1%	<div>Married</div>
NumCompaniesWorked	<div></div> 9.8%	<div>1</div>
JobLevel	<div></div> 9.54%	<div>1</div>
RelationshipSatisfaction	<div></div> 5.9%	<div>3</div>
PerformanceRating	<div></div> 5.17%	<div>3</div>
DistanceFromHome	<div></div> 3.22%	<div>2</div>
Gender	<div></div> 2.4%	<div>Male</div>
YearsAtCompany	<div></div> 1.93%	<div>5</div>

Attrition Prediction

No

Average prediction

No

55.8%

Yes

44.1%

Close

Download

AWS SAGEMAKER CANVAS

UPDATE THE OVERTIME AND CLICK UPDATE

employee-attrition-prediction

V1 Ready Add version Share

SelectBuildAnalyzePredict

Predict target values

Batch predictionSingle prediction

Modify values to predict Attrition in real time.

Filter columns

Column	Feature importance ↓	Value
EnvironmentSatisfaction	0.48%	3
HourlyRate	0.47%	66
MonthlyRate	0.44%	4223
EmployeeNumber	0.43%	1
OverTime	0.42%	Yes
Education	0.41%	3
BusinessTravel	0.34%	Travel_Rarely
StockOptionLevel	0.29%	3

Attrition Prediction

Copy

Yes

New prediction

Average prediction

No

36.2%

Yes

63.7%

CloseDownload

AWS SAGEMAKER CANVAS

YOU SHOULD SEE THE TRAINED MODEL
UNDER THE MODELS TAB

The screenshot displays the AWS SageMaker Canvas interface, specifically the 'Models' tab. The interface includes a sidebar with navigation icons, a top header with the 'Models' title and a '+ New model' button, and a main content area. In the main area, a model card for 'employee-attrition-prediction' is visible. The card shows a 'Ready' status with a green checkmark, a decorative graphic of two overlapping wavy bands (one purple, one blue), and a table of model details.

Property	Value
Accuracy	86.7%
Dataset	employee_data(2).csv
Target	Attrition
Problem type	2 category prediction

FINAL END-OF-DAY CAPSTONE PROJECT

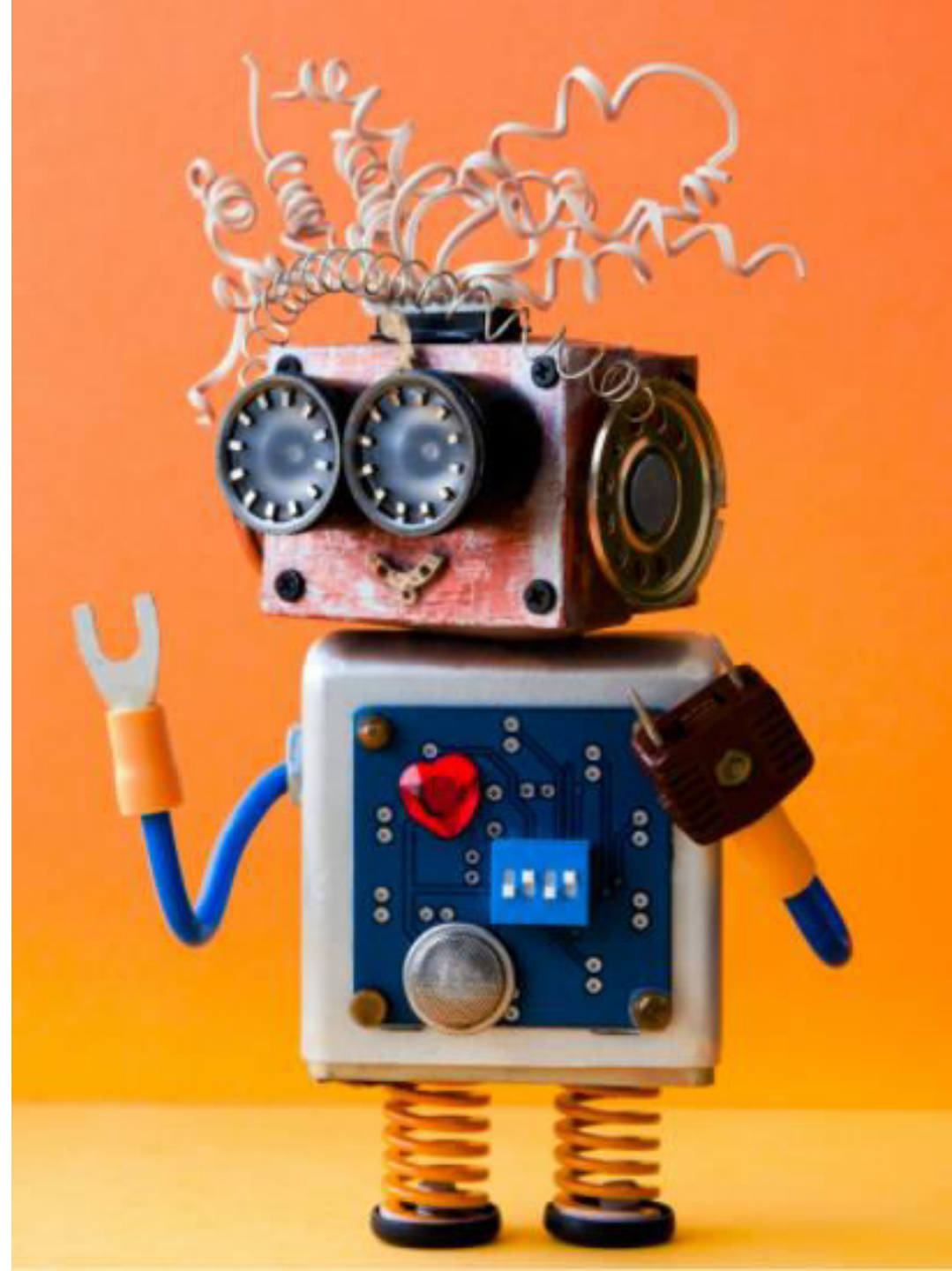


PROJECT

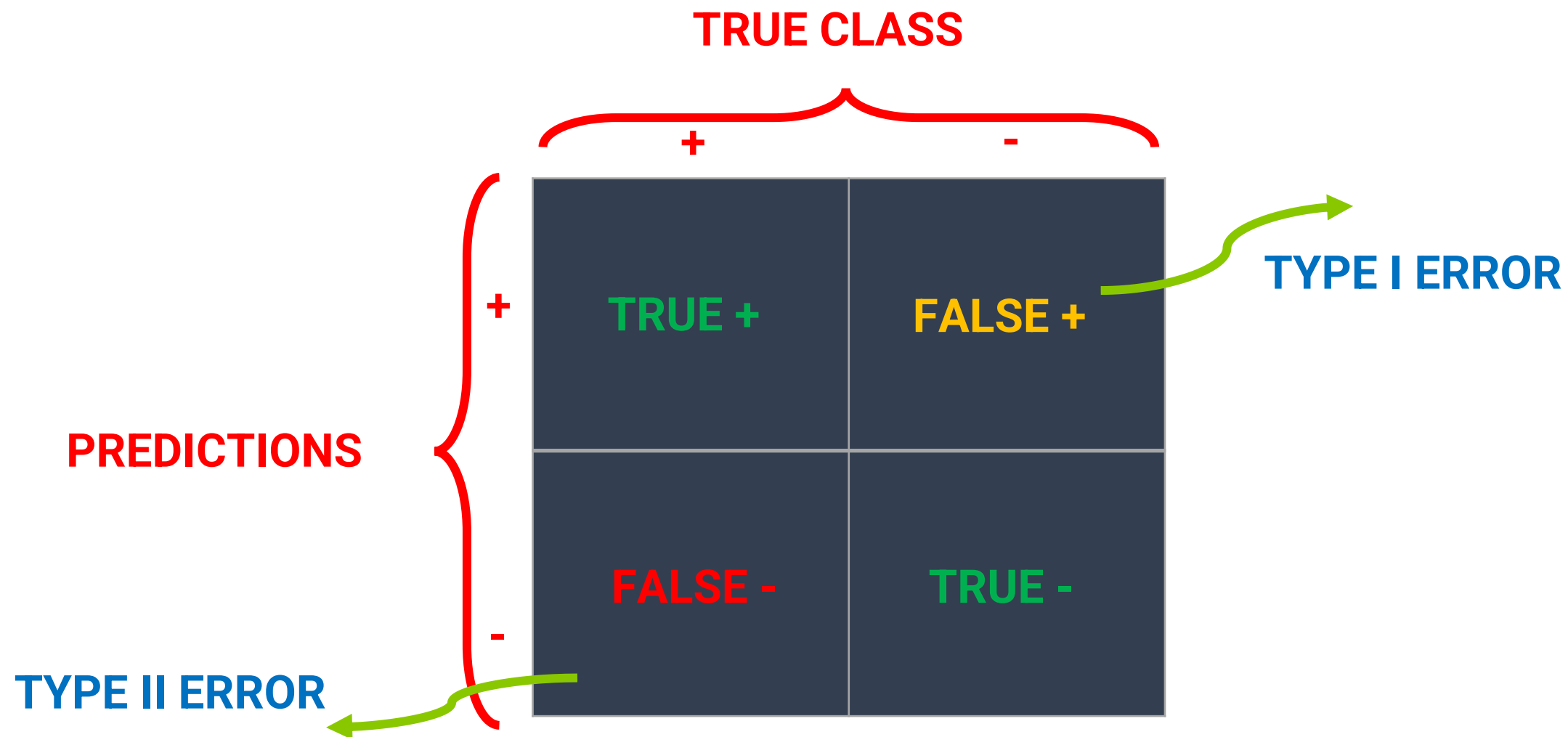
Using AWS SageMaker Canvas, perform the following tasks:

- 1. Upload the “*used_car_price.csv*” dataset to S3 (note that this datasets is a regression-type)
- 2. Perform basic data exploration
- 3. What is the average horsepower and its correlation to target?
- 4. Select the target column (MSRP)
- 5. Click on Preview model and list the estimated RMSE in \$ values
- 6. List the 3 most important features in the datasets
- 7. Train a regression model using quick build option
- 8. Test the model using at least 3 single prediction values

CLASSIFIER MODEL KEY PERFORMANCE INDICATORS (KPIs) – REVIEW



CONFUSION MATRIX

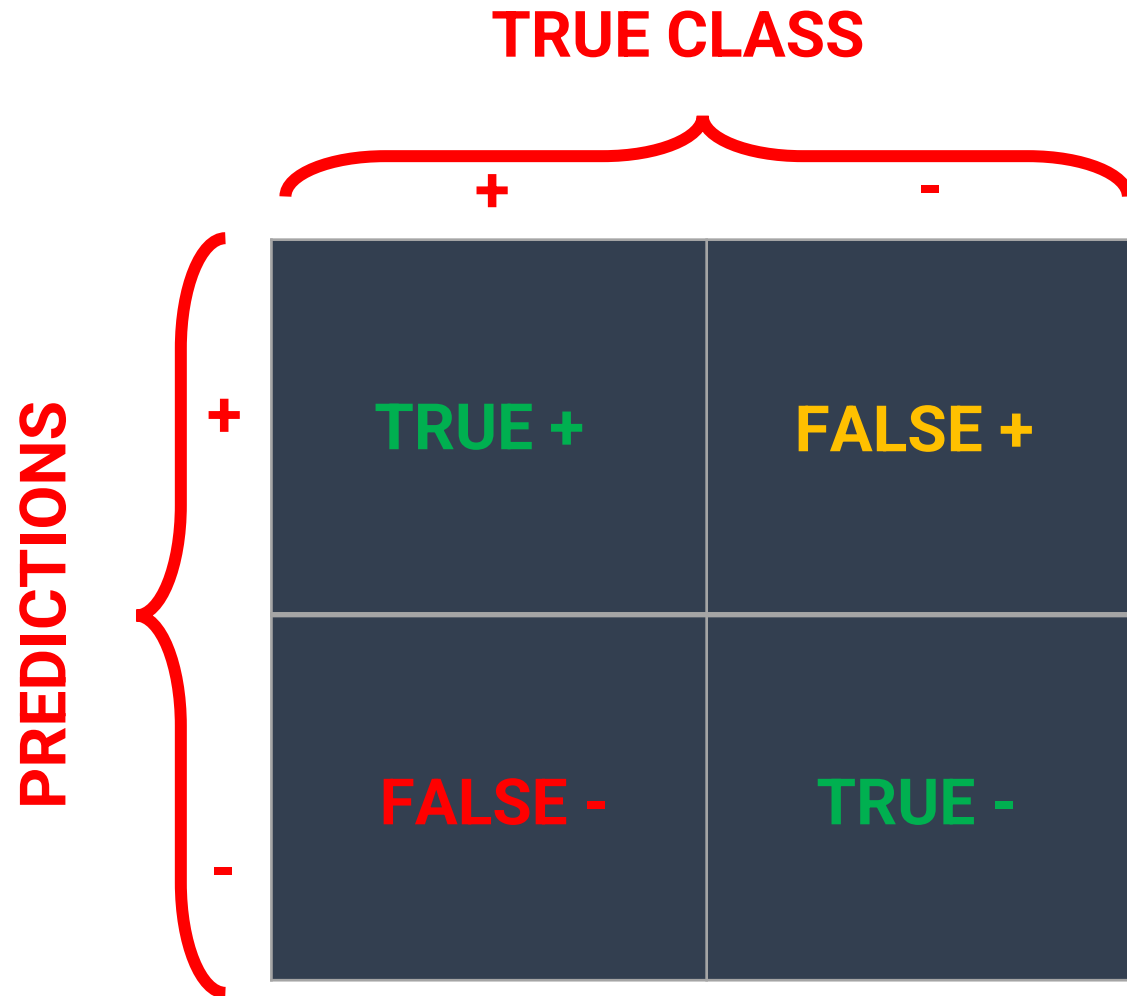


CLASSIFICATION MODEL KPIs

- A confusion matrix is used to describe the performance of a classification model:
 - True positives (TP): cases when classifier predicted TRUE (they have the disease), and correct class was TRUE (patient has disease).
 - True negatives (TN): cases when model predicted FALSE (no disease), and correct class was FALSE (patient do not have disease).
 - False positives (FP) (Type I error): classifier predicted TRUE, but correct class was FALSE (patient did not have disease).
 - False negatives (FN) (Type II error): classifier predicted FALSE (patient do not have disease), but they actually do have the disease
 - Classification Accuracy = $(TP+TN) / (TP + TN + FP + FN)$
 - Misclassification rate (Error Rate) = $(FP + FN) / (TP + TN + FP + FN)$

PRECISION Vs. RECALL

- Precision = $TP / \text{Total TRUE Predictions} = TP / (TP + FP)$ (When model predicted TRUE class, how often was it right?)
- Recall = $TP / \text{Actual TRUE} = TP / (TP + FN)$ (when the class was actually TRUE, how often did the classifier get it right?)



PRECISION Vs. RECALL

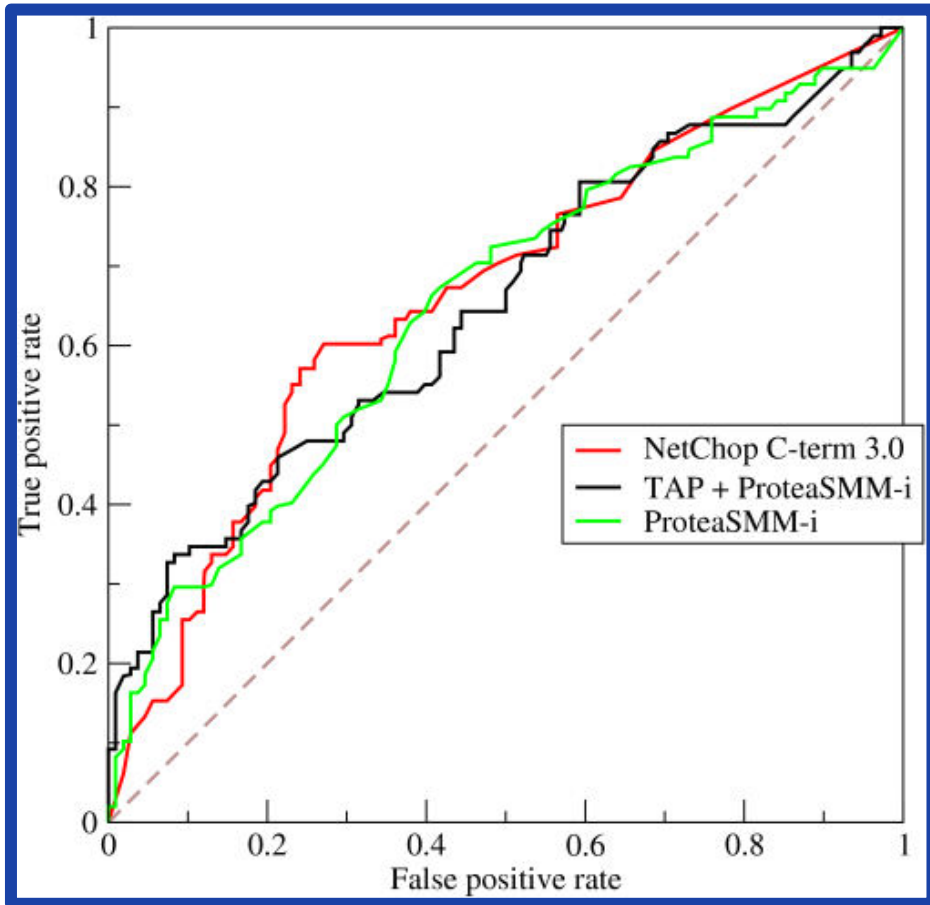
		TRUE CLASS	
		+	-
PREDICTIONS	+	TP = 1	FP = 1
	-	FN = 8	TN = 90

FACTS:
100 PATIENTS TOTAL
91 PATIENTS ARE HEALTHY
9 PATIENTS HAVE CANCER

- Accuracy is generally misleading and is not enough to assess the performance of a classifier.
- Recall is an important KPI in situations where:
 - Dataset is highly unbalanced; cases when you have small cancer patients compared to healthy ones.

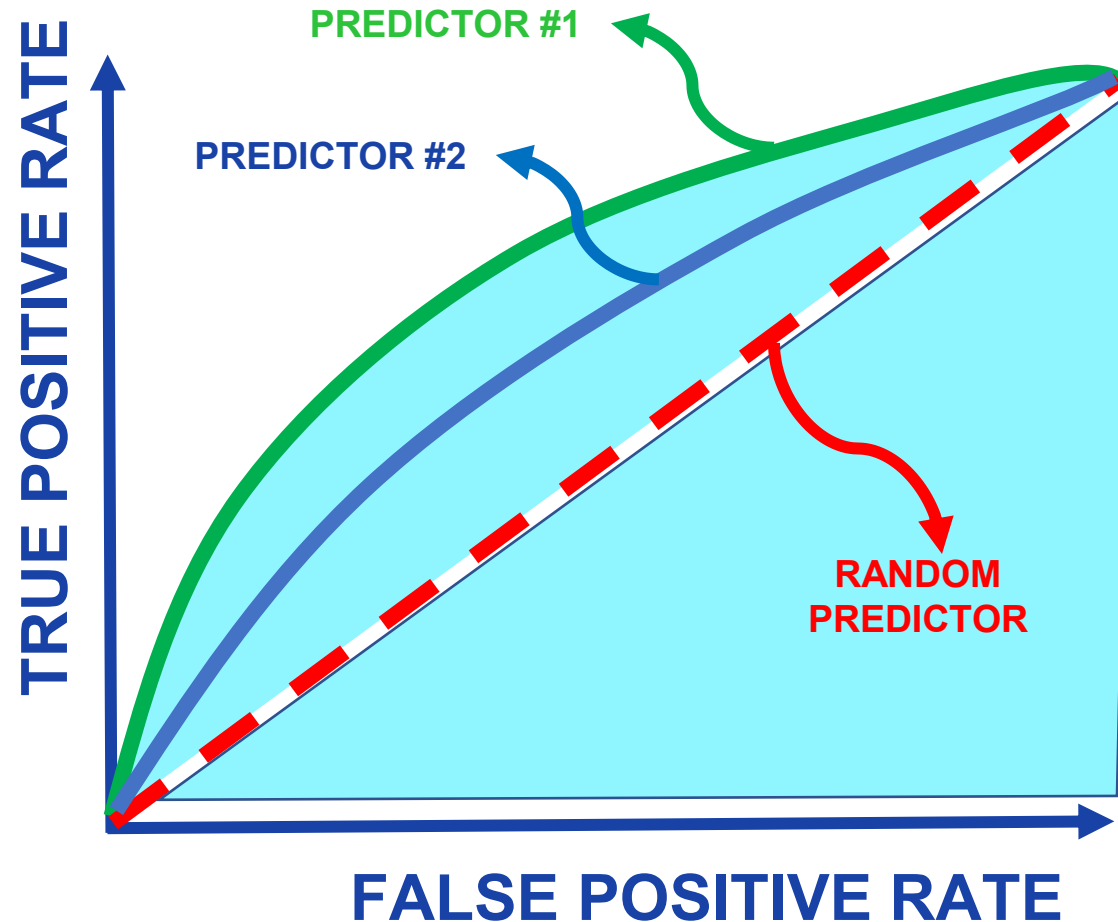
- Classification Accuracy = $(TP+TN) / (TP + TN + FP + FN) = 91\%$
- Precision = $TP/\text{Total TRUE Predictions} = TP / (TP+FP) = 1/2 = 50\%$
- Recall = $TP/\text{Actual TRUE} = TP / (TP+FN) = 1/9 = 11\%$

ROC (RECEIVER OPERATING CHARACTERISTIC CURVE)



- ROC Curve is a metric that assesses the model ability to distinguish between binary (0 or 1) classes.
- The ROC curve is created by plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings.
- The true-positive rate is also known as sensitivity, recall or probability of detection in machine learning.
- The false-positive rate is also known as the probability of false alarm.
- Points above the diagonal line represent good classification (better than random)
- The model performance improves if it becomes skewed towards the upper left corner.

AUC (AREA UNDER CURVE)



- The light blue area represents the area Under the Curve of the Receiver Operating Characteristic (AUROC).
- The diagonal dashed red line represents the ROC curve of a random predictor with AUROC of 0.5.
- If ROC AUC = 1, perfect classifier
- Predictor #1 is better than predictor #2
- Higher the AUC, the better the model is at predicting 0s as 0s and 1s as 1s.