

Data Science Workshop

An Experiential Journey with Data to Inspire *Your* Work

George Stark

Maureen Norton

Neeraj Madan

January 29, 2020

The Open Group Conference
Making Data Useful
San Antonio, Texas
January 29, 2020
9:00 – 12:30

Data Science Workshop

An Experiential Journey with Data to Inspire *Your* Work

January 27, 2020

January 28, 2020

January 29, 2020

January 30, 2020

Data Science Workshop

9:00 AM - 12:30 PM

An Experiential Journey With Data to Inspire Your Work

The Experiential Journey with Data to Inspire Your Work session will make you think differently about data and how it can solve problems! You will hear surprising use cases that will make you think, sometimes laugh and hopefully inspire your own work. The use cases and introductory material will be followed by a hands-on experiential journey described below. The most valuable part of this session is that it is designed to help you gain experience and relate it to your work – so that when you leave you have a plan of action on how you can make data more useful in your organization to solve a key challenge.

A real-business application of analytics in “Improving Customer Experiences with Real-Time Insights” will be used as an example during the workshop. This experiential session will include a step by step journey on “How data science is helping IBM to predict the customer experience journey and proactively address the issues, leading to the improvement of Net Promoter Score”. The session will also highlight the importance of using CRISP-DM (Cross Industry Standard Process for Data Mining) and Agile in Data Science projects.

The methodology involves consuming historical NPS data; using machine learning and artificial intelligence to identify the most important features and created an algorithm to predict the customer experience.

Facilitators: [Neeraj Madan](#), [Maureen Norton](#), [George Stark](#)

The Open Group Conference
Making Data Useful
San Antonio, Texas
January 29, 2020
9:00 – 12:30

Agenda

Section	Time
Getting Started a. Session Introduction and Expectation Setting b. Data Science Introduction <i>a. Let's Talk About Data</i> <i>b. Common Business Models</i> <i>c. Data Science Techniques</i> <i>d. Participant Workbook</i>	9:00 am – 10:00 am
Predictive Analytics and Machine Learning Solutions IBM Watson Studio and Machine Learning (Introduction and Setup) << Break >>	10:00 am – 10:40 am
Hands on Experiential Journey (Net Promoter Score Example) a. Business understanding: <i>Exercise 1: Identify an opportunity in your business context and document</i> b. Data understanding: <i>Exercise 2: What data set would you gather to work the problem statement</i> c. Data preparation: <i>Exercise 3: How would you prepare the dataset and what challenges do you foresee?</i> <<Break (15 mins)>> d. Modeling: <i>Exercise 4: What modeling techniques would you attempt and why?</i> e. Evaluation: <i>Exercise 5: What metrics would you use to evaluate your model performance?</i> f. Deployment: <i>Exercise 6: How do you plan to consume the outputs of the model?</i>	10: 45 am – 12:30 pm

Let's talk about data

Is there a source of data that has information about

- ANY topic
- ANY where
- ANY time

Is there a source that has their finger on the pulse of what people think at any moment in time?



Twitter

Soggy Fries



Let's talk about data

What other types
of data can be
used to drive
deeper insights?

WEATHER

The image features a dramatic sky scene. The bottom half is filled with large, white, fluffy clouds. Above the clouds is a bright blue horizon line. The top half of the image is a dark blue sky with faint, thin, curved lines that resemble orbits or weather patterns. The word "WEATHER" is written in large, white, bold, sans-serif capital letters across the middle of the image, partially overlapping the clouds and the blue sky.

Four Common Data Science Models

Risk Assessment

Create a “Screening Model” to identify “threats”. Threats can be any sort of fraudulent activity (e.g., credit transaction, passenger screening, ability to purchase, altered video/photo, Fake/Real news)



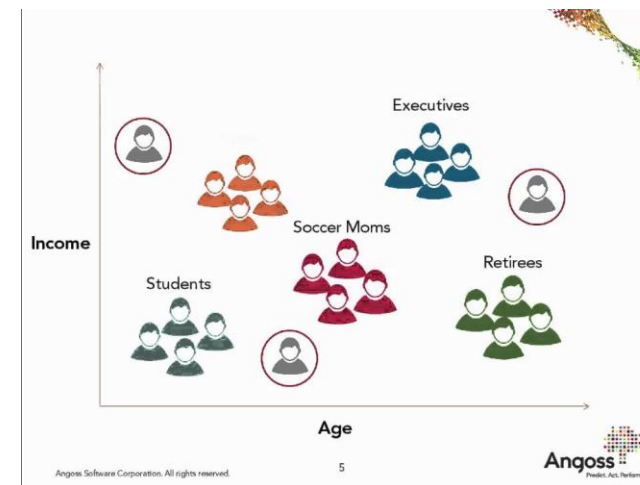
Quality/Defect Prediction

Identify problematic components, predict number of defects in a product (e.g., code, castings, compounds, raw materials, ATM Machines)



Business Value/ Customer Satisfaction

Create a classification algorithm that accurately identifies which customers have the most potential business value based on their characteristics and activities. Which customers are likely to be happy? Which will be promoters?

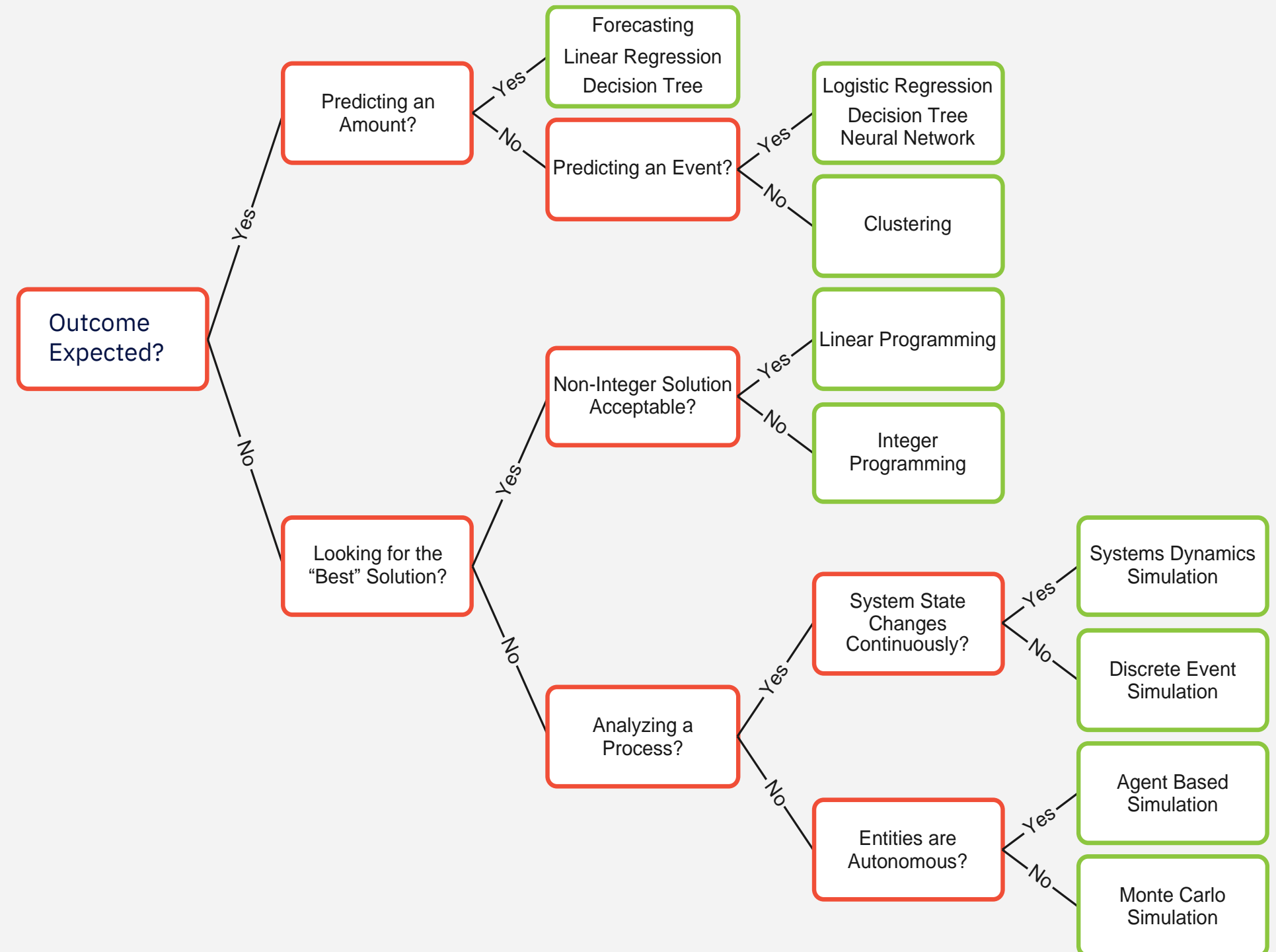


Price/Cost/Value

Predict value (e.g., home/ rental prices, value of retail transaction, number of issues, etc)



Choosing the right analytic approach



Data Science Project One-Pager

Project Description					Business Lead: Last, First Data Science Lead: Last, First Architect: Last, First
Business Understanding	Data Understanding	Data Preparation	Modeling	Evaluation	Deployment
<ul style="list-style-type: none"> ✓ Current Business Situation ✓ Decision Maker Identification ✓ Data-Mining Success Criteria from a Business Perspective <ul style="list-style-type: none"> ✓ <u>Cupcake Deliverable</u>: "Classify a transaction with class of profit" ✓ <u>Wedding Cake Deliverable</u>: "Predict with 80% Confidence Interval (CI) the GP of a transaction" 	<ul style="list-style-type: none"> ❑ Existing Data Sources <ul style="list-style-type: none"> ✓ Customer Loyalty ❑ Sales ❑ Users (via API) ❑ Financial ❑ SKU ❑ Additional Data <ul style="list-style-type: none"> ❑ MagicHat 	<ul style="list-style-type: none"> ❑ Merging ❑ Feature Selection ❑ Aggregation ❑ Feature Engineering <ul style="list-style-type: none"> ❑ OneHotEncoding ❑ TF-IDF ❑ Principal Component Analysis (PCA) ❑ Data Quality Strategy <ul style="list-style-type: none"> ❑ Drop ❑ Imputation <ul style="list-style-type: none"> ❑ Mean ❑ Median ❑ ML-based (KNN) ❑ Test/Train Split <ul style="list-style-type: none"> ❑ K-fold CV 	<ul style="list-style-type: none"> ❑ Model Types <ul style="list-style-type: none"> ❑ Random Forest Regressor ❑ KNN-based ensemble ❑ Generalized Additive Models (GAM) ❑ Model Requirements <ul style="list-style-type: none"> ❑ L1/L2 Regularization ❑ Ensemble Pipeline ❑ LIME-interpolation 	<ul style="list-style-type: none"> ❑ Hyper parameter Validation <ul style="list-style-type: none"> ❑ Validation Curve ❑ GridSearch CV ❑ Visualizations ❑ Evaluation Metrics <ul style="list-style-type: none"> ❑ Classification <ul style="list-style-type: none"> ❑ Accuracy ❑ Precision ❑ Recall ❑ Confusion Matrix ❑ Regression <ul style="list-style-type: none"> ❑ R2 ❑ RMSE ❑ Business Feedback 	<ul style="list-style-type: none"> ❑ Summary Findings ❑ Integration ❑ Model Monitoring ❑ Model Maintenance ❑ Enhancements <ul style="list-style-type: none"> ❑ Recommendations ❑ UI/UX Design Thinking ❑ Cognitive

Predictive Analytics and Machine Learning Solutions

To name a few:

1. IBM Watson Studio
2. SAS Advanced Analytics
3. RapidMiner
4. Amazon SageMaker
5. Azure Machine Learning Studio (Microsoft)
6. Google Cloud AI Platform

“The Forrester Wave™: Multimodal Predictive Analytics And Machine Learning Solutions, Q3 2018”, Forrester Research, September 2018

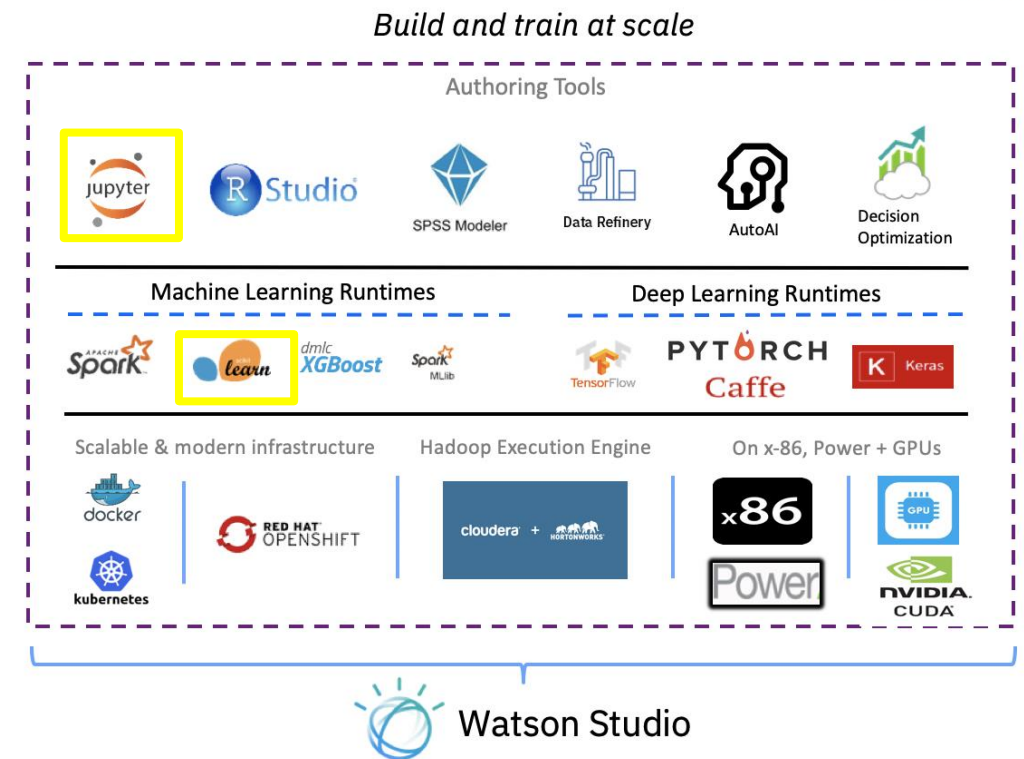
IBM Watson Studio

(Enterprise Data Science platform that helps your team work together to build models to make better data driven decisions for your business)

Analyze any data, no matter where it lives

Empower your entire organization with notebooks, visual productivity, and automation tools

One platform to rule them all from discovery to production



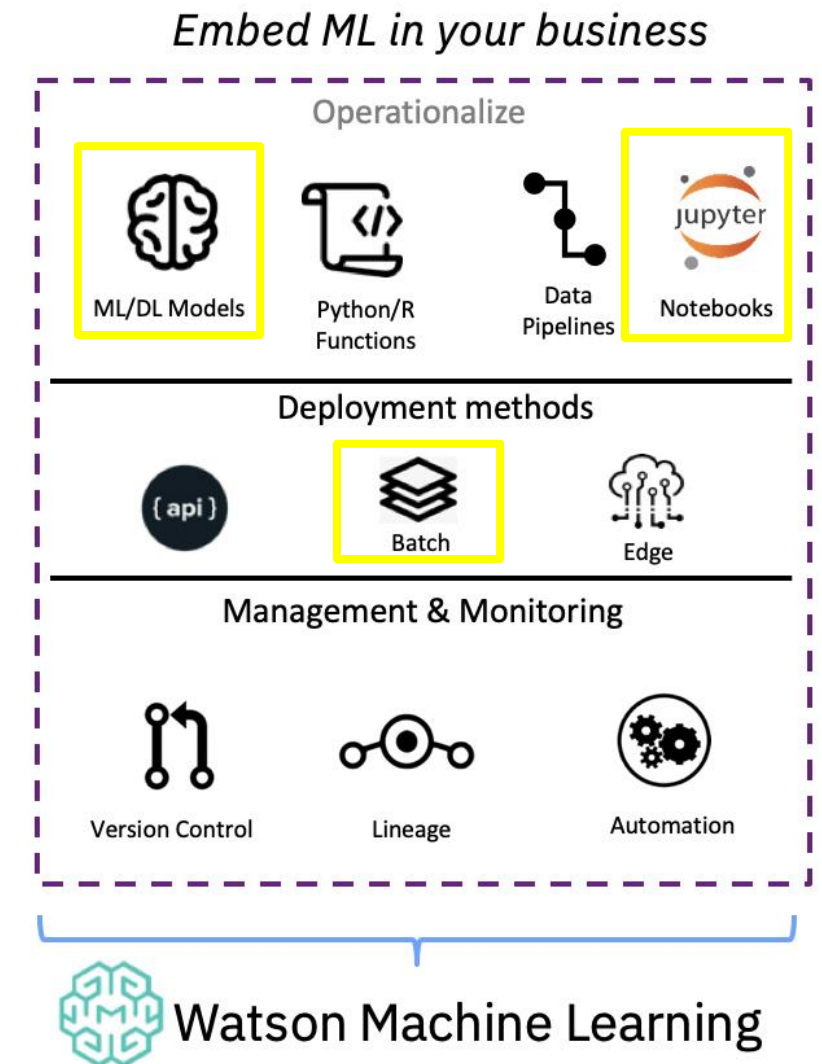
IBM Watson Machine Learning

(Embed Machine Learning and Deep Learning in your Business)

Deploy and Manage Models

Intelligent Model Operations

Accelerate Compute Intensive Workloads



Getting Started: IBM Watson Studio Setup

(Step 1: Sign up / Log into IBM Cloud - <https://ibm.biz/BdqDmG>)

Already have an IBM Cloud account? [Log in](#)

Create a free account

Join us in the cloud and start building today.

Email

Enter an email address.

First Name

Last Name

Country or Region

United States

Password

IBM may use my contact data to keep me informed of products, services and offerings:

☒ by email. ☒ by telephone.

You can withdraw your marketing consent at any time by sending an email to netsupp@us.ibm.com. Also you may unsubscribe from receiving marketing emails by clicking the unsubscribe link in each such email.

More information on our processing can be found in the [IBM Privacy Statement](#). By submitting this form, I acknowledge that I have read and understand the IBM Privacy Statement.

I accept the product [Terms and Conditions](#) of this registration form.

Create Account

© Copyright IBM Corp. 2014, 2019. All rights reserved.



Already have an IBM Cloud account? [Log in](#)

Create a free account

Join us in the cloud and start building today.

Email

First Name

Neeraj

Last Name

Madan

Country or Region

United States

Password

IBM may use my contact data to keep me informed of products, services and offerings:

☒ by email. ☒ by telephone.

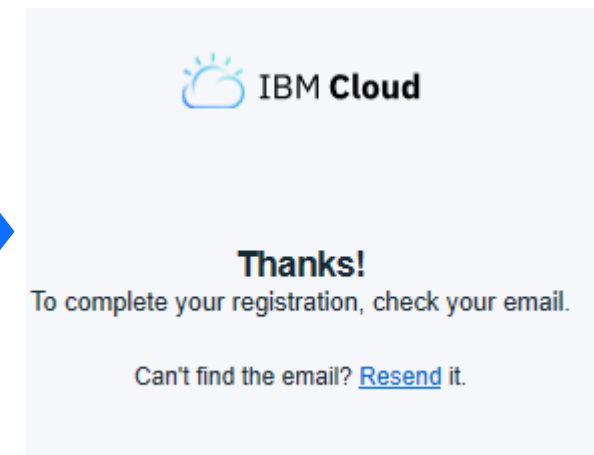
You can withdraw your marketing consent at any time by sending an email to netsupp@us.ibm.com. Also you may unsubscribe from receiving marketing emails by clicking the unsubscribe link in each such email.

More information on our processing can be found in the [IBM Privacy Statement](#). By submitting this form, I acknowledge that I have read and understand the IBM Privacy Statement.

I accept the product [Terms and Conditions](#) of this registration form.

Create Account

© Copyright IBM Corp. 2014, 2019. All rights reserved.



Getting Started: IBM Watson Studio Setup

(Step 2: Confirm your account (Email verification))

Action required: Confirm your IBM Cloud account



Hello Neeraj,

Thank you for signing up for IBM Cloud! Confirm your account to get started.

[Confirm account](#)

By confirming your account, you accept the [Terms of Use](#).

Welcome and happy building!

Thank you,
IBM Cloud

Visit the [IBM Cloud console](#).

© Copyright IBM Corporation 2014, 2020.

IBM



Welcome!

You confirmed your IBM Cloud account, and it is now activated. Log in to get started.

[Log in](#)

About your IBMid Account Privacy

This notice provides information about accessing your IBMid user account (Account). If you have previously been presented with a version of this notice, please refer to "Changes since the previous version of this notice" below for information about the new updates.

[+ Changes since the previous version of this notice](#)

[+ What data does IBM collect?](#)

[+ Why IBM needs your data](#)

[+ How your data was obtained](#)

[+ How IBM uses your data](#)

[+ How IBM protects your data](#)

[+ How long we keep your data](#)

Your rights

Our [Privacy Statement](#) provides more information about your personal data rights. It also provides contact information if you have questions or concerns regarding our handling of your personal data.

Acknowledgement

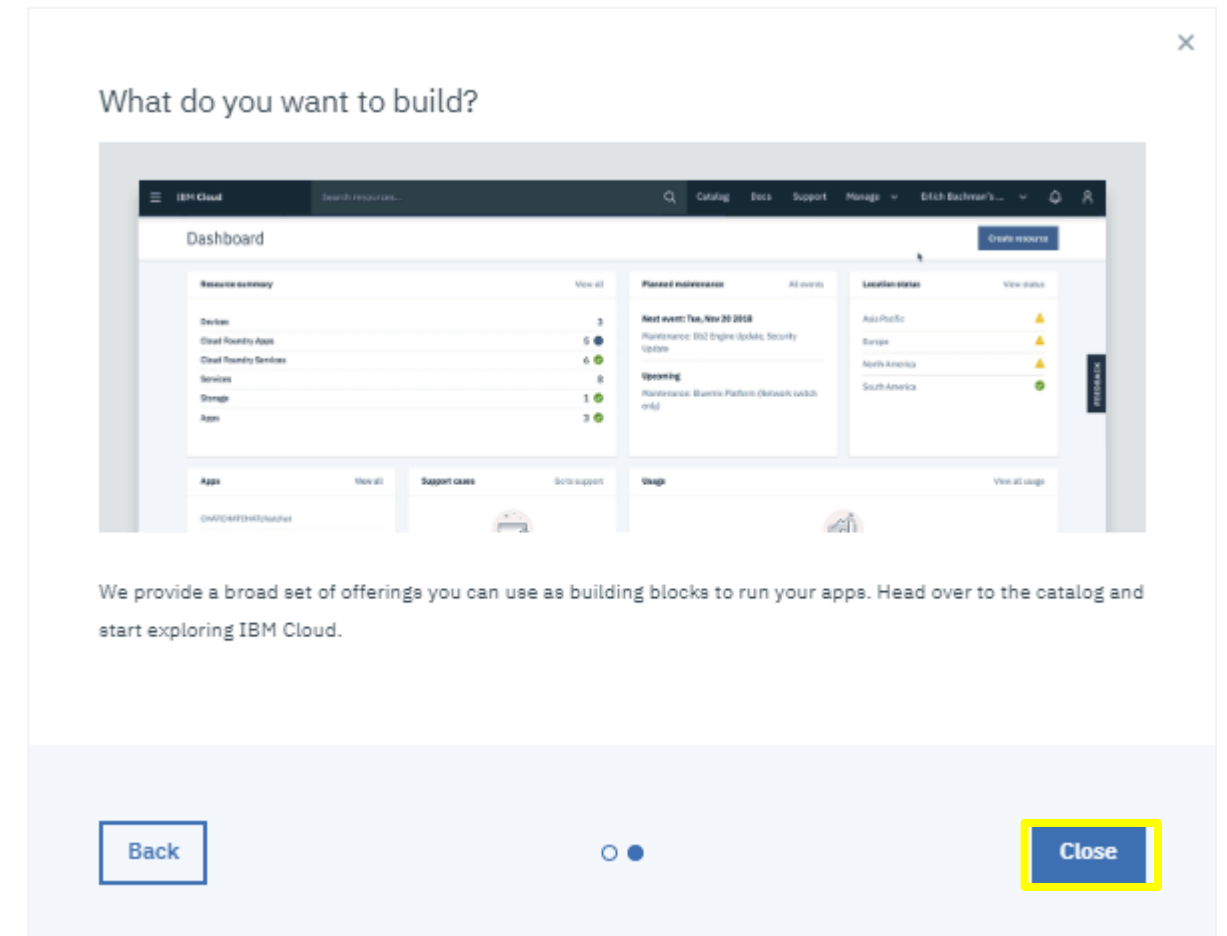
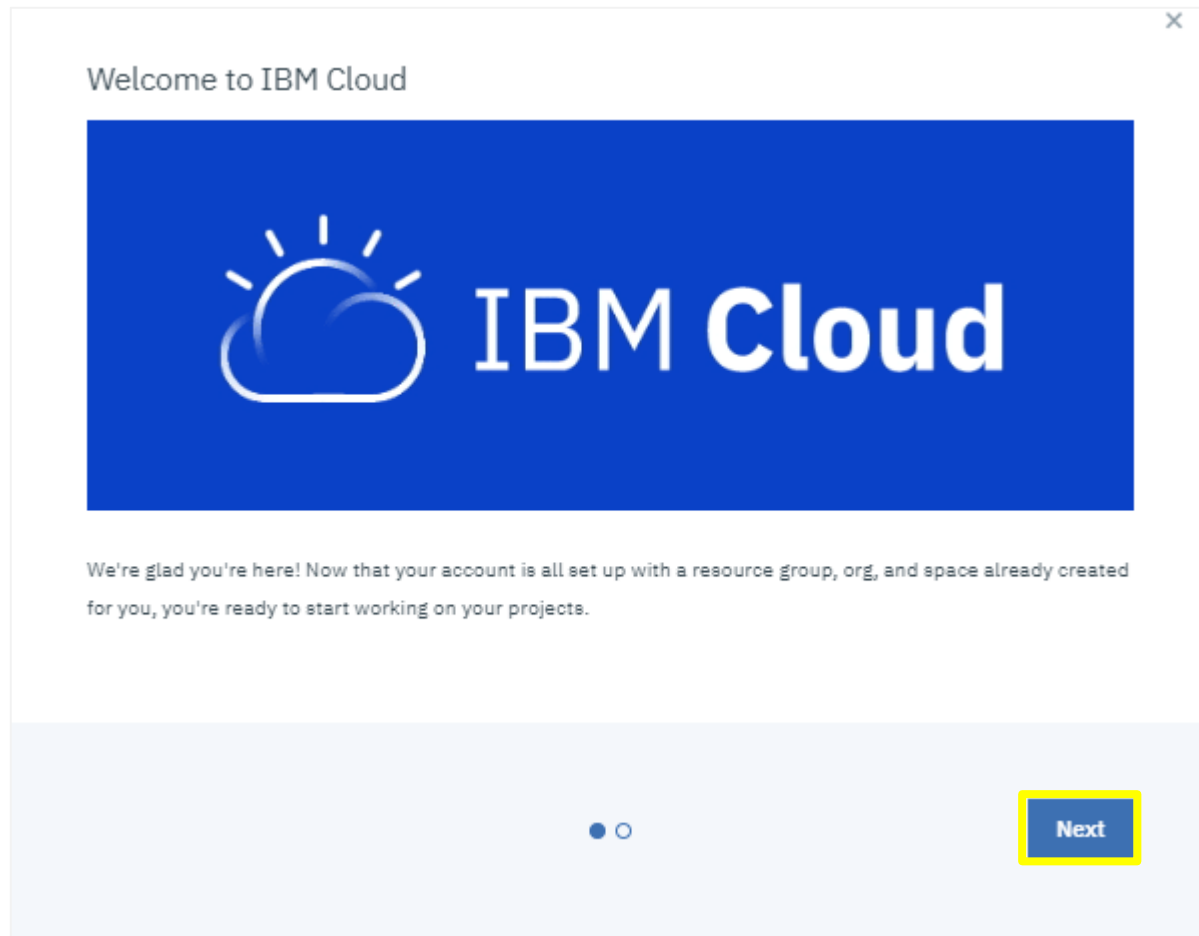
I acknowledge that I understand how IBM is using my Basic Personal Data and I am at least 16 years of age.

[Proceed](#)

[Cancel Sign In](#)

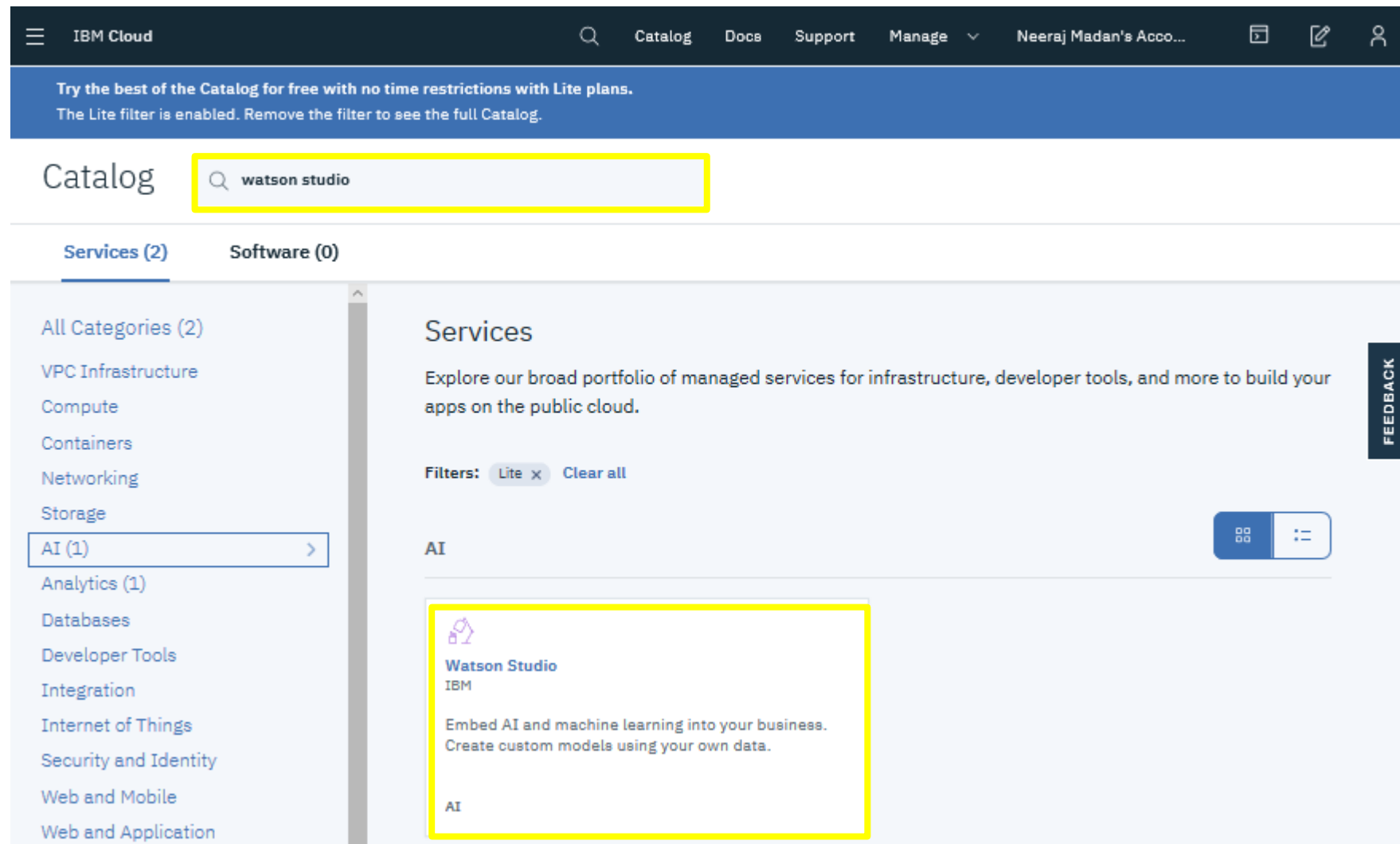
This document was last updated on 2018-05-04

Getting Started: IBM Watson Studio Setup (Step 3: Welcome to IBM Cloud)



Getting Started: IBM Watson Studio Setup

(Step 4: Locate Watson Studio in the Catalog)



Getting Started: IBM Watson Studio Setup (Step 5: Create a Watson Studio Service)

IBM Cloud

Watson Studio Lite IBM Service IAM-enabled

Author: IBM • Date of last update: 01/13/2020

Need Help? [Contact Support](#) [View docs](#)

Create About

Select a region

Dallas

Select a pricing plan

Displayed prices do not include tax. Monthly prices shown are for country or region: [United States](#)

PLAN	FEATURES	PRICING
✓ Lite	1 authorized user 60 capacity unit-hours monthly limit 1 free small compute environment with 1 vCPU and 4 GB RAM (does not require capacity unit-hours)	Free

The Lite plan for Watson Studio offers everything you need to become a better data scientist or domain expert in a collaborative environment.

Lite plan services are deleted after 30 days of inactivity.

Summary

Watson Studio *Free*

Region: Dallas

Plan: Lite

Service name: Watson Studio-79

Resource group: Default

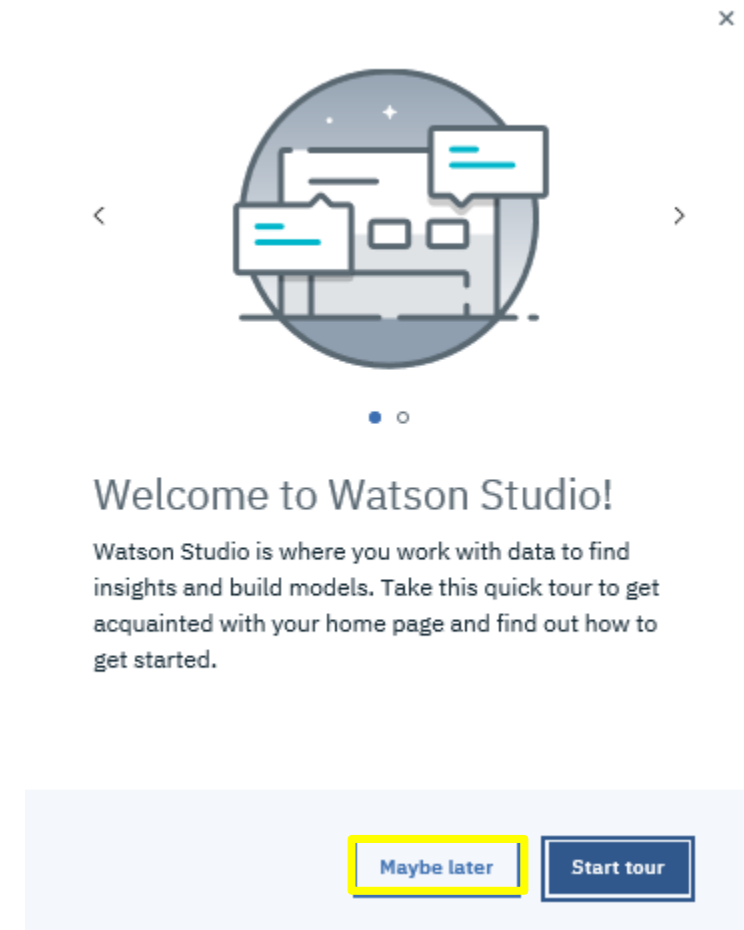
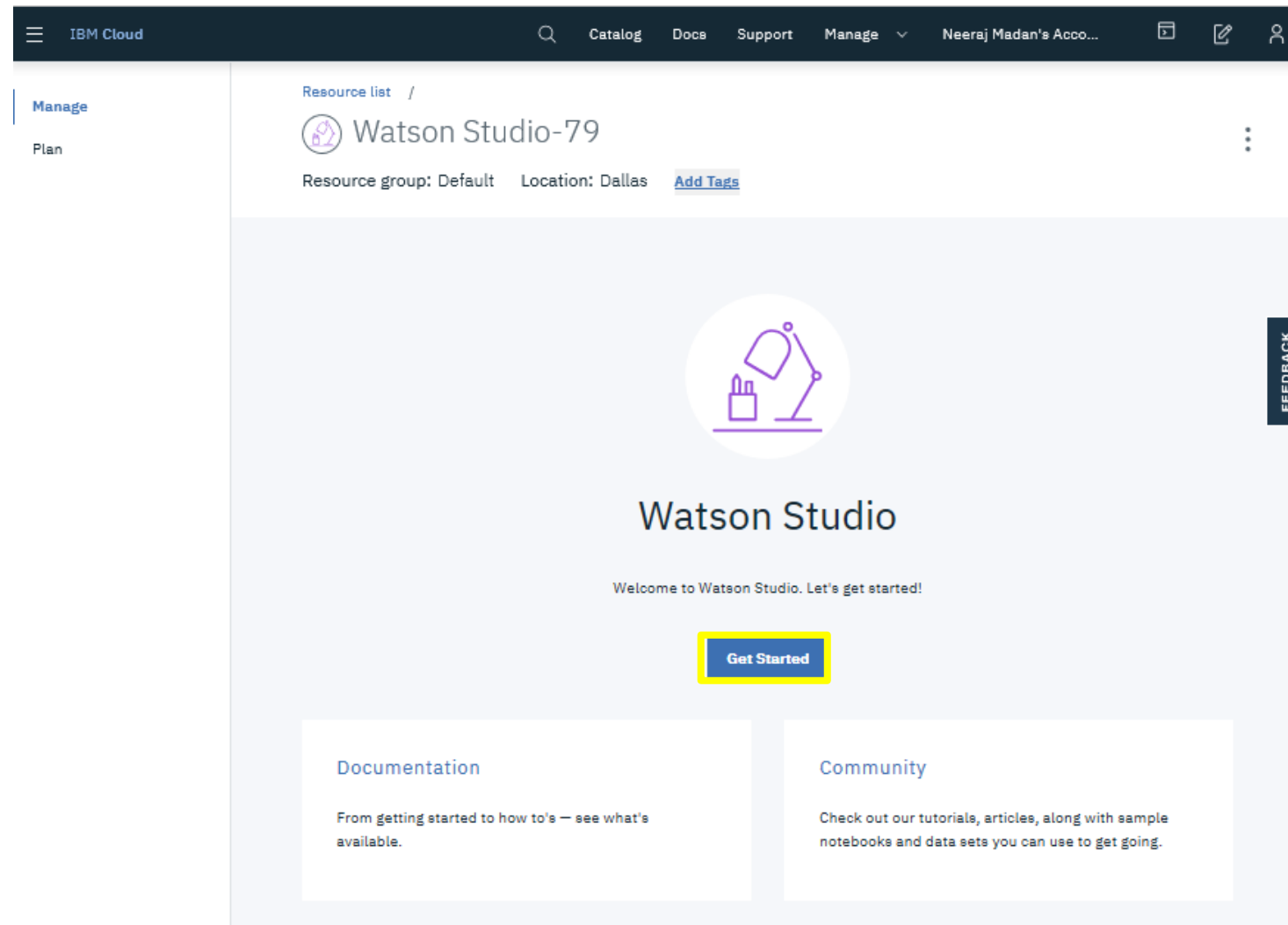
Create

Add to estimate

[View terms](#)

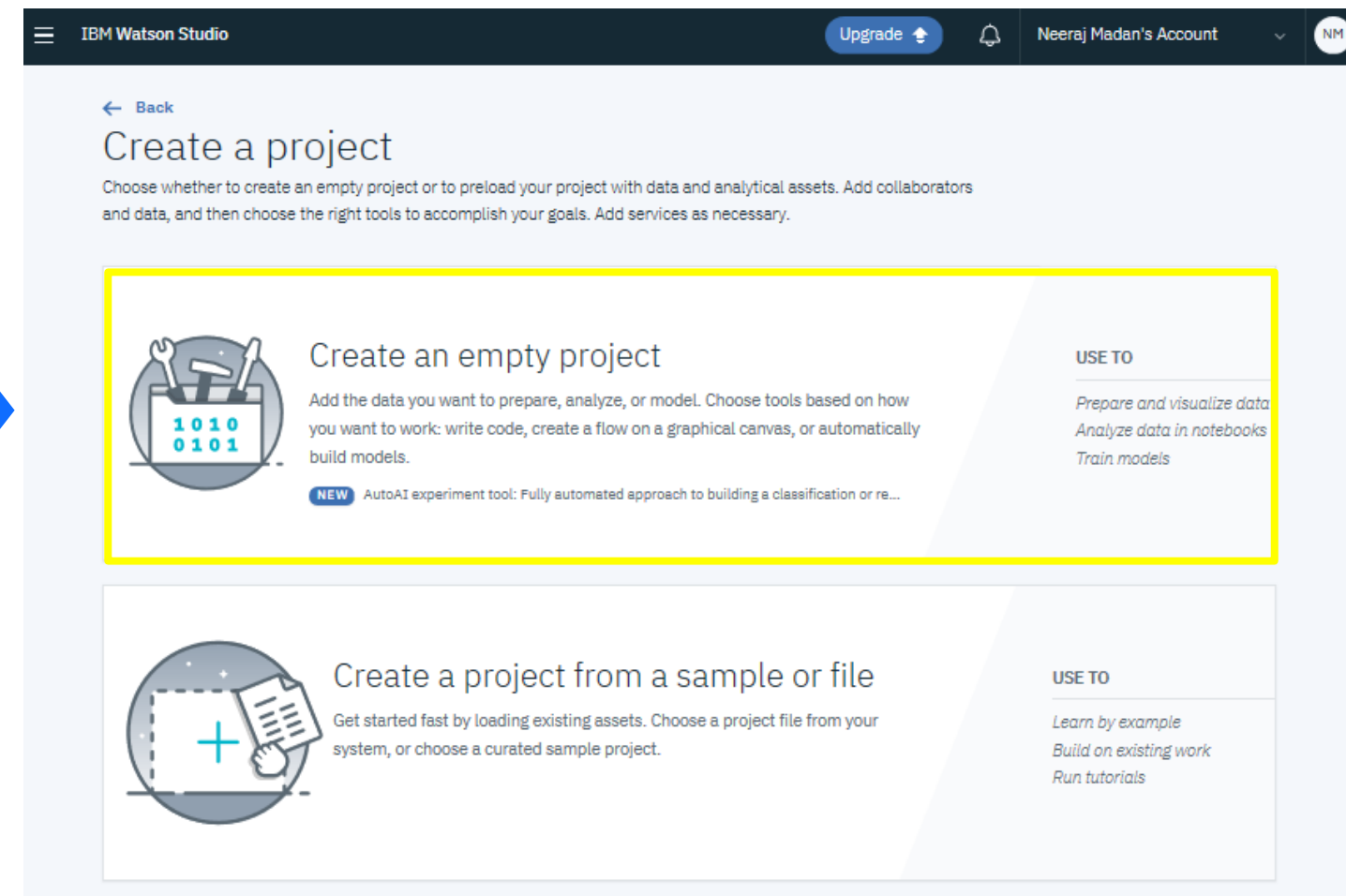
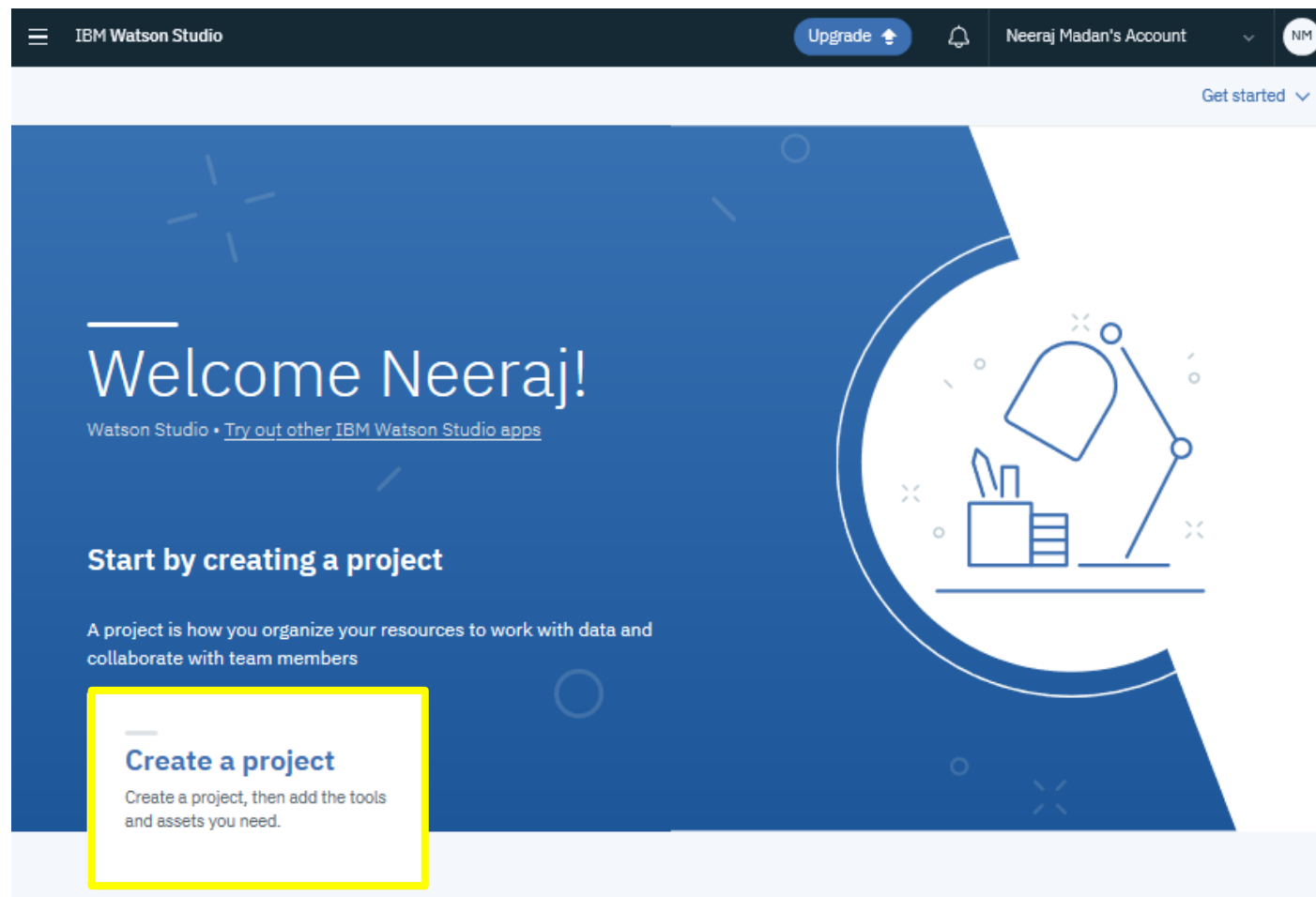
FEEDBACK

Getting Started: IBM Watson Studio Setup (Step 6: Launch Watson Studio)








Getting Started: IBM Watson Studio Setup

(Step 7: Create a project and pick an empty template)



Getting Started: IBM Watson Studio Setup

(Step 8a: Project and Storage Setup)

 IBM Watson Studio Upgrade   Neeraj Madan's Account  

New project

Define project details


Name

NPS Prediction

Description

Project description

Choose project options

☐ Restrict who can be a collaborator 

Project will include integration with [Cloud Object Storage](#) for storing project assets.

Define storage

① Select storage service

Add

Add an object storage instance and then return to this page and click Refresh.

② Refresh

Getting Started: IBM Watson Studio Setup

(Step 8b: Project and Storage Setup)

IBM Watson Studio

Upgrade

Neeraj Madan's Account

NM

Existing

New

Cloud Object Storage

IBM Cloud Object Storage is a highly scalable cloud storage service, designed for high durability, resiliency and security. Store, manage and access your data via our self-service portal and RESTful APIs. Connect applications directly to Cloud Object Storage use other IBM Cloud Services with your data.

Features

Storage for the IBM Cloud

IBM Cloud Object Storage provides unstructured data storage for cloud applications. Libraries and SDKs support a common set of S3 API functions for connecting new applications to scalable cloud storage and integrating your data into other services on the IBM Cloud Platform as well as IBM Watson services. IBM Cloud Object Storage is available with Regional, Cross Region and single site resiliency options worldwide.

Built-in Aspera high-speed transfer

With IBM Cloud Object Storage Aspera high-speed data transfer, you can improve data transfer performance by quickly transferring data over long distances, and under various network conditions. It is natively integrated into Cloud Object Storage and there is no additional cost for uploading data.

Storage Classes and Archive Policy

Choose storage classes based on your usage patterns for active, less-active, and cold workloads with Standard, Vault, and Cold Vault respectively. Use Flex class for dynamic data access with usage patterns that are hard to predict. For rarely used data that requires long-term retention, simply set an Archive policy with our existing storage-class tiers allowing you to reduce costs even further with our lowest priced Archive storage.

Access and Key Management

IBM Identity and Access Management (IAM) policies allow for granular access control at the bucket level using role-based policies. Key Protect support allows customers to have their own managed encryption keys for higher level data security.

Pricing Plan: Monthly Process shown above reflect the: **United States**

PLAN	FEATURES	PRICING
<input checked="" type="radio"/> Lite	<p>1 COS Service Instance</p> <p>Storage up to 25 GB/mo.</p> <p>Up to 20,000 GET requests/mo.</p> <p>Up to 2,000 PUT requests/mo.</p> <p>Up to Data Retrieval 10 GB/mo.</p> <p>Up to 5GB Public Outbound</p> <p>Applies to aggregate total across all storage bucket classes</p>	Free
<input type="radio"/> Standard	There is no minimum fee, so you pay only for what you use.	Expand each section to view details

The Lite service plan for Cloud Object Storage includes Regional and Cross Regional resiliency, flexible data classes, and built in security.

Cancel

Create



Confirm Creation

Plan

Lite

Resource group

Default

Service name

cloud-object-storage-qx

Cancel

Confirm



Getting Started: IBM Watson Studio Setup (Step 8c: Project and Storage Setup)

IBM Watson Studio

Upgrade

Neeraj Madan's Account

NM

New project

Define project details

Name

NPS Prediction

Description

Project description

Choose project options

☐ Restrict who can be a collaborator

Project will include integration with Cloud Object Storage for storing project assets.

Define storage

1 Select storage service

Add

Add an object storage instance and then return to this page and click Refresh.

2 Refresh

Storage

cloud-object-storage-qx

Cancel Create



Gather your resources

Your project can contain many types of resources, but the most important are your team and your data. Add collaborators and data sets, and then create analytic assets like notebooks and models to work with your data.

Maybe later Start tour

Getting Started: IBM Watson Studio Setup

(Step 9a: Create a project token)

The screenshot illustrates the steps to create a project token in IBM Watson Studio. It starts with the main dashboard for the 'NPS Prediction' project. The 'Settings' tab is highlighted in the top navigation bar. A blue arrow points down to the 'Access tokens' section, where a 'New token' button is highlighted. Another blue arrow points down to the 'New Token' modal form. In this form, the 'Name' field is set to 'NPS Prediction' and the 'Access role for project' is set to 'Editor'. The 'Create' button is highlighted at the bottom right of the modal.

IBM Watson Studio

Upgrade

Neeraj Madan's Account

My Projects / NPS Prediction

Launch IDE

Add to project

Overview Assets Environments Jobs Deployments Access Control **Settings**

Access tokens

New token

NAME	ROLE	CREATED	LAST USED	ACTIONS
You don't have any Access tokens yet.				

New Token

Name

NPS Prediction

Access role for project

Editor

Cancel Create

Getting Started: IBM Watson Machine Learning Setup

(Step 9b: Add Watson Machine Learning Service)

Associated services

NAME	SERVICE TYPE	PLAN	ACTIONS
You don't have any Associated services yet.			

Add service

- Amazon EMR Spark
- IBM Analytics Engine
- Streaming Analytics
- Dashboard
- Watson**

IBM Watson Studio Upgrade Neeraj Madan's Account NM

Discovery
Add a cognitive search and content analytics engine to applications.
[Add](#)

Language Translator
Translate text, documents, and websites from one language to another. Create industry or region-specific models.
[Add](#)

Machine Learning
IBM Watson Machine Learning - make smarter decisions, solve tough problems, and improve user outcomes.
[Add](#)

Natural Language Classifier
Natural Language Classifier uses advanced natural language processing and machine learning techniques to create custom models.
[Add](#)

Natural Language Understanding
Analyze text to extract meta-data from content such as concepts, entities, emotion, relations, sentiment.
[Add](#)

Personality Insights
The Watson Personality Insights derives insights from transactional and social media data to identify psychological traits.
[Add](#)

Speech to Text
Low-latency, streaming transcription.
[Add](#)

Text to Speech
Synthesizes natural-sounding speech from text.
[Add](#)

Tone Analyzer
Tone Analyzer uses linguistic analysis to detect three types of tones from communications: emotion, sentiment, and style.
[Add](#)

Visual Recognition
Find meaning in visual content! Analyze images for scenes, objects, and other content. Choose a default model.
[Add](#)

Getting Started: IBM Watson Machine Learning Setup

(Step 9c: Add Watson Machine Learning Service)

IBM Watson Studio

Upgrade

Neeraj Madan's Account

NM

Machine Learning

IBM Watson Machine Learning is a full-service IBM Cloud offering that makes it easy for developers and data scientists to work together to integrate predictive capabilities with their applications. The Machine Learning service is a set of REST APIs that you can call from any programming language to develop applications that make smarter decisions, solve tough problems, and improve user outcomes.

Features

Machine Learning features

Take advantage of machine learning models management (continuous learning system) and deployment (online, batch, streaming). Select any of widely supported machine learning frameworks: TensorFlow, Keras, Caffe, PyTorch, Spark MLlib, scikit learn, xgboost and SPSS.

Wide choice of interfaces

Use the command line interface and Python client to manage your artifacts. Extend your application with artificial intelligence through the Watson Machine Learning REST API.

Integration with Watson Studio

Create and train machine learning models with the best tools and the latest expertise in a social environment built by and for data scientists.

Pricing Plan: Monthly Process shown above reflect the: United States

PLAN	FEATURES	PRICING
<div>Lite</div>	<div>Service instance</div> <div>5 model deployments</div> <div>5,000 predictions</div> <div>50 capacity unit-hours (CUH) included:</div> <div>Capacity Type:</div> <div><ul style="list-style-type: none">1 (one) NVIDIA K80 GPU = 2 capacity units required per hour1 (one) NVIDIA V100 GPU = 8 capacity units required per hour1 vCPU and 4 GB RAM = 0.5 capacity units required per hour2 vCPU and 8 GB RAM = 1 capacity units required per hour4 vCPU and 16 GB RAM = 2 capacity units required per hour8 vCPU and 32 GB RAM = 4 capacity units required per hour16 vCPU and 64 GB RAM = 8 capacity units required per hour</div> <div>Auto AI</div> <div>8 vCPU and 32 GB RAM = 20 capacity units required per hour</div> <div>Decision Optimization</div> <div>2 vCPU and 8 GB RAM = 30 capacity units required per hour</div> <div>4 vCPU and 16 GB RAM = 40 capacity units required per hour</div> <div>16 vCPU and 64 GB RAM = 60 capacity units required per hour</div>	Free

The lite plan instance of the IBM Watson Machine Learning service provides you with a maximum of 5 deployed models, 5,000 predictions per month, and 50 capacity unit-hours per month during which model can be trained, evaluated, and deployed to be available to accept prediction events, with a minimum of 1 minute per training job.

Create



Confirm Creation

Region

Dallas

Plan

Lite

Resource group

Default

Service name

pm-20-uu

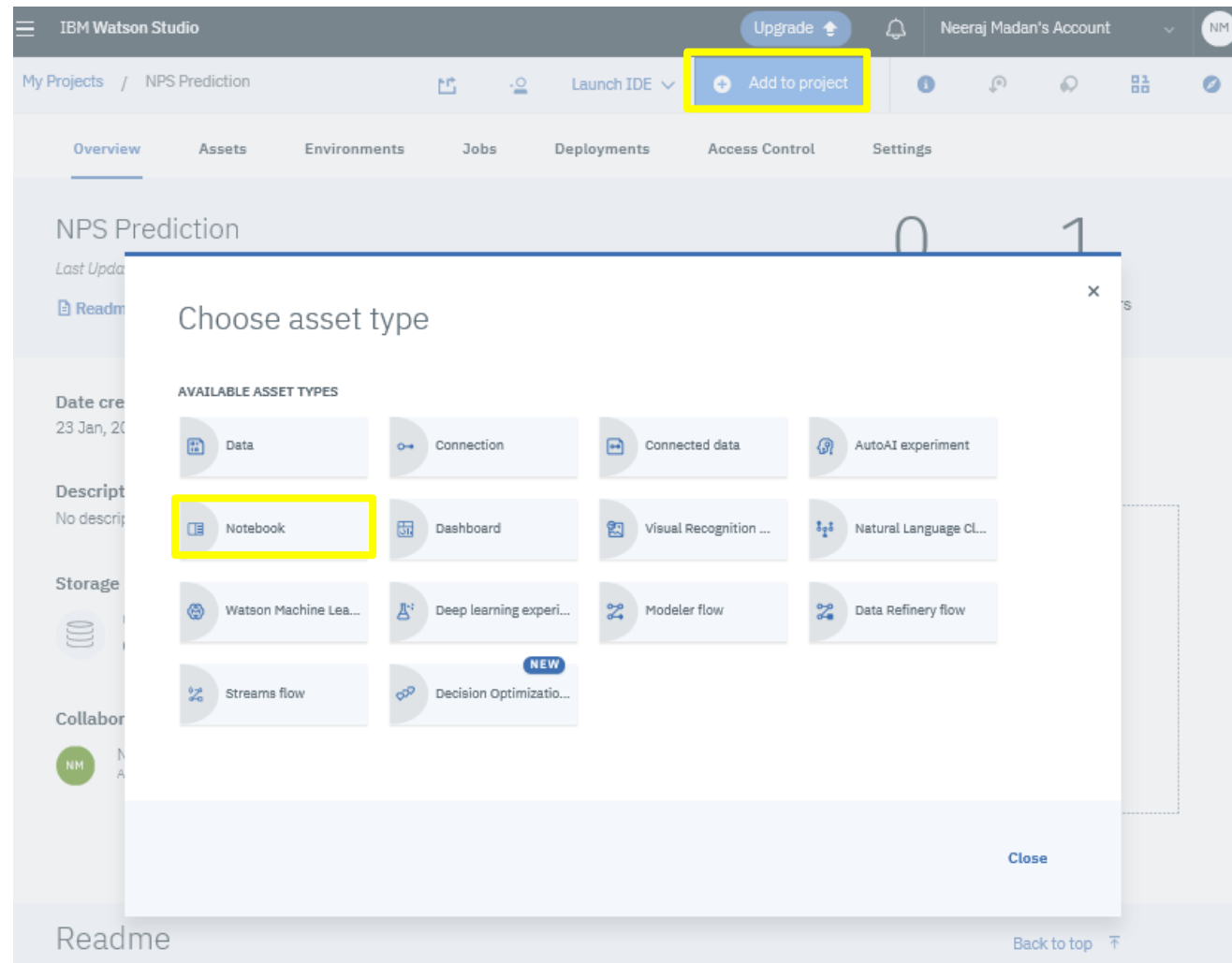
Cancel

Confirm



Getting Started: Load Notebook

(Step 10a: Add a notebook to your project)



Getting Started: Load Notebook

(Step 10b: Import a notebook, get the link from github)

IBM Watson Studio

Upgrade

Neeraj Madan's Account

My Projects / NPS Prediction / Add Notebook

New notebook

Blank From file From URL

Name

NPS Prediction

26 characters remaining

Description (optional)

Type your Description here

500 characters remaining

Select runtime

Default Python 3.6 Free (1 vCPU and 4 GB RAM)

The selected runtime has 1 vCPU and 4 GB RAM and is free.
[Learn more](#) about capacity unit hours and Watson Studio pricing plans.

Notebook URL

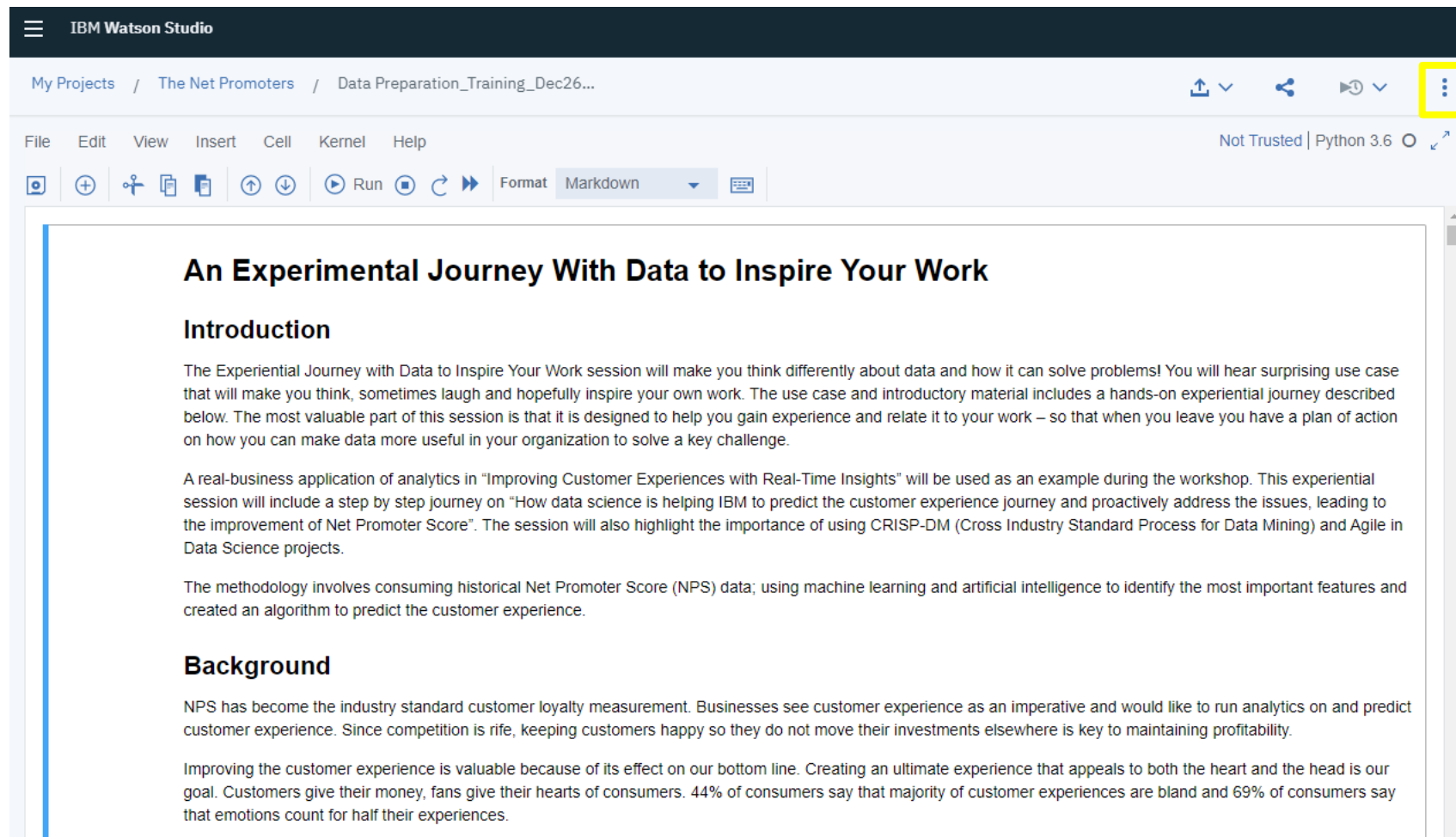
http://bit.ly/ogsat_dsw

Link: http://bit.ly/ogsat_dsw

GitHub Link: http://bit.ly/OGSAT2020_GitHub

Getting Started: Load Notebook

(Step 10c: Let's get started!)



The screenshot displays the IBM Watson Studio web interface. At the top, a dark blue header bar contains the 'IBM Watson Studio' logo. Below this, a breadcrumb trail shows the path: 'My Projects / The Net Promoters / Data Preparation_Training_Dec26...'. To the right of the breadcrumb, there are icons for upload, share, and a menu (three vertical dots, highlighted with a yellow box). Below the header, a light blue toolbar contains various icons for file operations (add, delete, copy, paste, move) and a 'Run' button. To the right of the toolbar, the text 'Not Trusted | Python 3.6' is displayed. The main content area is a white notebook with a blue vertical margin on the left. The notebook title is 'An Experimental Journey With Data to Inspire Your Work'. Below the title is the section 'Introduction'. The text in the introduction describes an experiential journey with data to inspire work, mentioning a hands-on session and the use of CRISP-DM and Agile in Data Science projects. Below the introduction is the section 'Background'. The text in the background describes NPS as an industry standard customer loyalty measurement and the goal of improving customer experience.

IBM Watson Studio

My Projects / The Net Promoters / Data Preparation_Training_Dec26...

File Edit View Insert Cell Kernel Help

Not Trusted | Python 3.6

An Experimental Journey With Data to Inspire Your Work

Introduction

The Experiential Journey with Data to Inspire Your Work session will make you think differently about data and how it can solve problems! You will hear surprising use case that will make you think, sometimes laugh and hopefully inspire your own work. The use case and introductory material includes a hands-on experiential journey described below. The most valuable part of this session is that it is designed to help you gain experience and relate it to your work – so that when you leave you have a plan of action on how you can make data more useful in your organization to solve a key challenge.

A real-business application of analytics in “Improving Customer Experiences with Real-Time Insights” will be used as an example during the workshop. This experiential session will include a step by step journey on “How data science is helping IBM to predict the customer experience journey and proactively address the issues, leading to the improvement of Net Promoter Score”. The session will also highlight the importance of using CRISP-DM (Cross Industry Standard Process for Data Mining) and Agile in Data Science projects.

The methodology involves consuming historical Net Promoter Score (NPS) data; using machine learning and artificial intelligence to identify the most important features and created an algorithm to predict the customer experience.

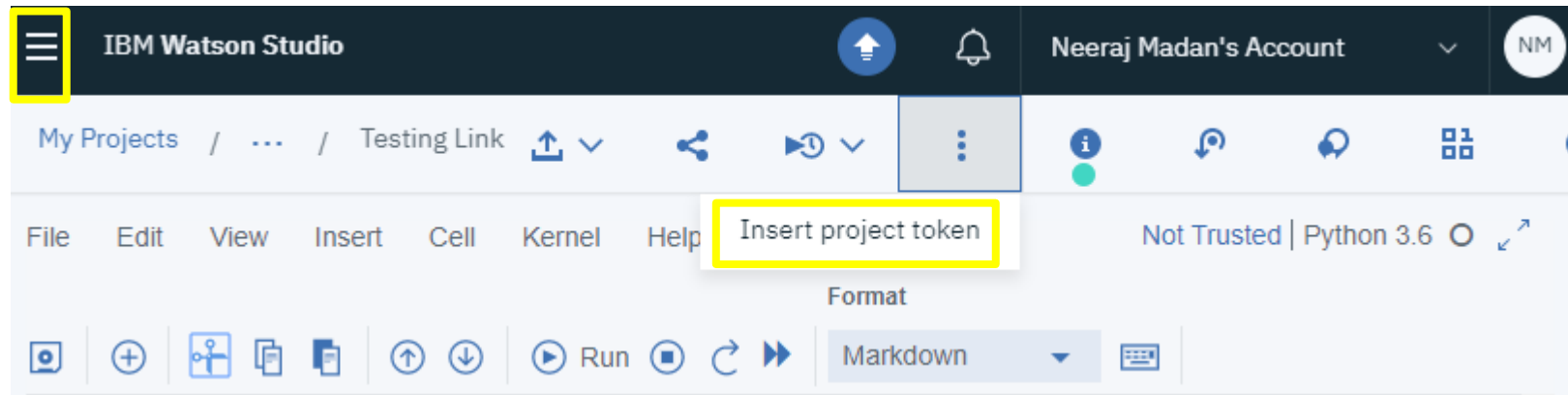
Background

NPS has become the industry standard customer loyalty measurement. Businesses see customer experience as an imperative and would like to run analytics on and predict customer experience. Since competition is rife, keeping customers happy so they do not move their investments elsewhere is key to maintaining profitability.

Improving the customer experience is valuable because of its effect on our bottom line. Creating an ultimate experience that appeals to both the heart and the head is our goal. Customers give their money, fans give their hearts of consumers. 44% of consumers say that majority of customer experiences are bland and 69% of consumers say that emotions count for half their experiences.

Getting Started: Load Notebook

(11. Update your credentials in the Notebook)



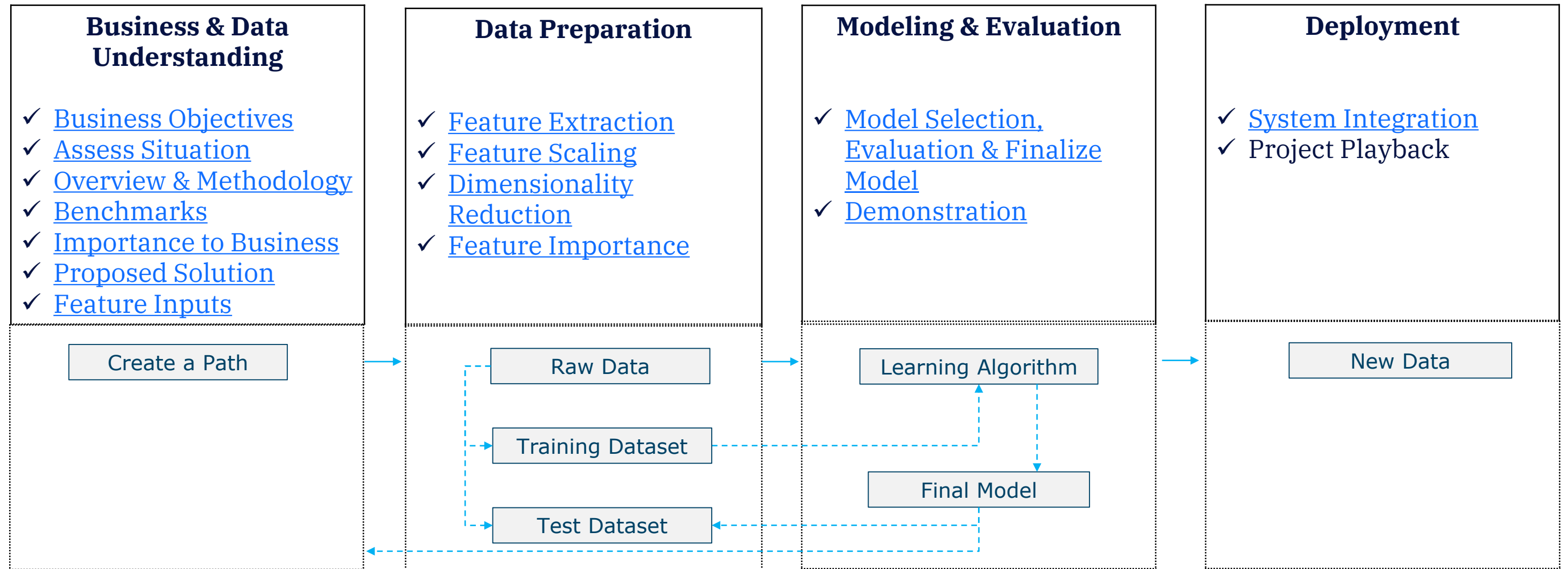
Getting Started: Load Notebook (12. Machine Learning Credentials)

Keep these credentials handy as we are going to use them while running the notebook

```
{  "apikey": "",  "iam_apikey_description": "Auto-generated for key ",  "iam_apikey_name": "Service credentials-1",  "iam_role_crn": "crn:v1:bluemix:public:iam::::serviceRole:Writer",  "iam_serviceid_crn": "crn:v1:bluemix:public:iam-identity::",  "instance_id": "",  "url": "https://us-south.ml.cloud.ibm.com"}
```


Framework: Roadmap to Building Machine Learning System

"Essentially, all models are wrong, but some are useful."--- Box, George E. P.; Norman R. Draper (1987). Empirical Model-Building and Response Surfaces, p. 424, Wiley. ISBN 0471810339.



Take away

Now, I am able to

- ✓ create/ setup the Data Science environment on IBM Cloud
- ✓ learn/ re-use the roadmap to build a machine learning system

Business Objectives

Goal: Improve the Net Promoter Score by identifying potential non promoters ahead of time and proactively address customer issues.

Methodology: Consumed historical NPS data; Used machine learning and artificial intelligence to identify the most important features and created an algorithm to predict the non promoters.

Desired Result: To create an insert into production environment (ticketing system) to indicate top candidates for non-promoter surveys.

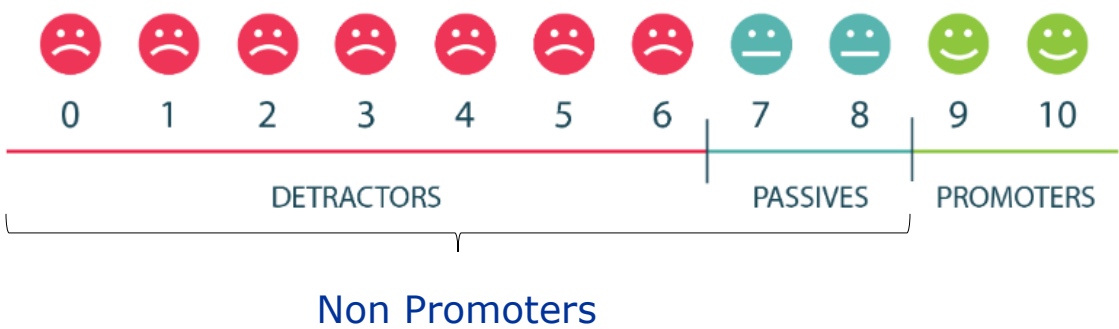


Assess Situation

In year 2018, IBM world-wide **supported 500,000 cases** which were created in multiple platforms.

The **Net Promoter Survey (NPS)** response rate was **15%**.

60% cases were non-promoters and 40% were promoters.



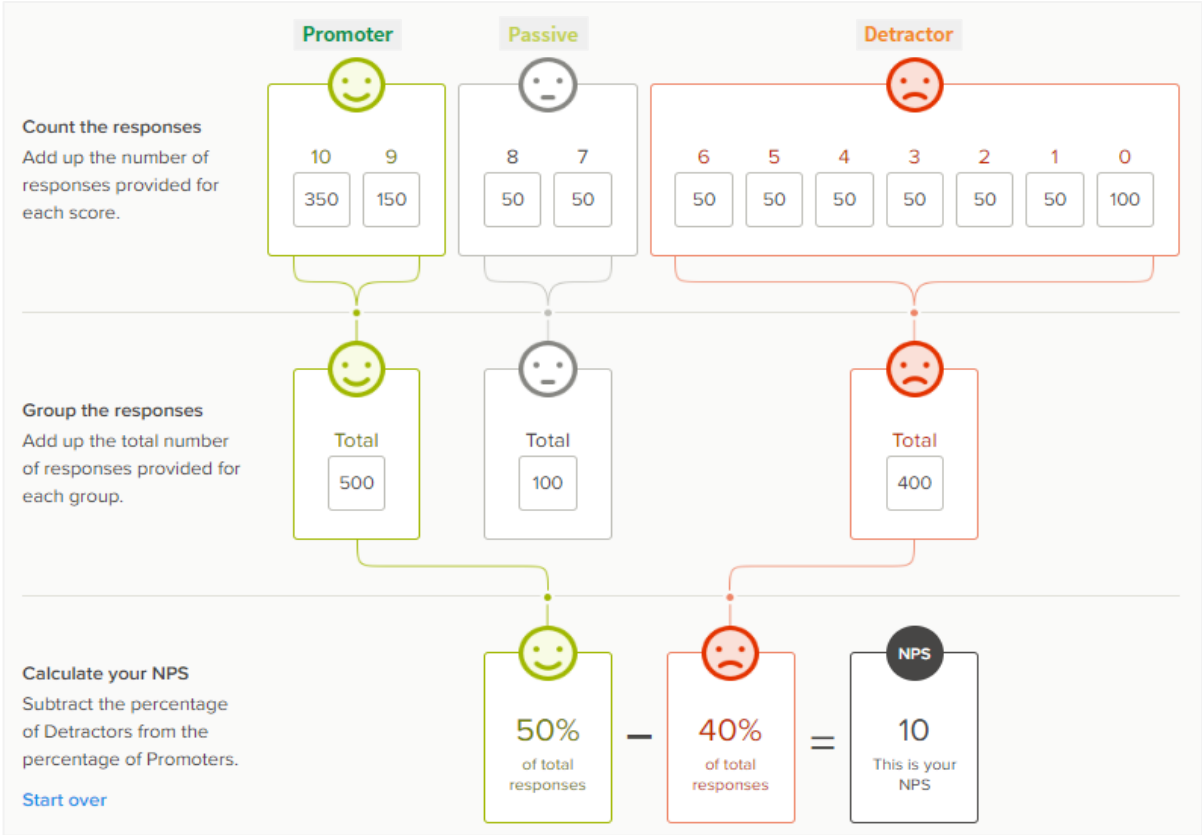
Note: The numbers highlighted above are crafted for this workshop.

Overview and Methodology

Net Promoter has become the industry standard customer loyalty measurement. Businesses see customer experience as an imperative.

On a scale of 0-10, how likely would you recommend [brand/ support] to a friend or colleague?

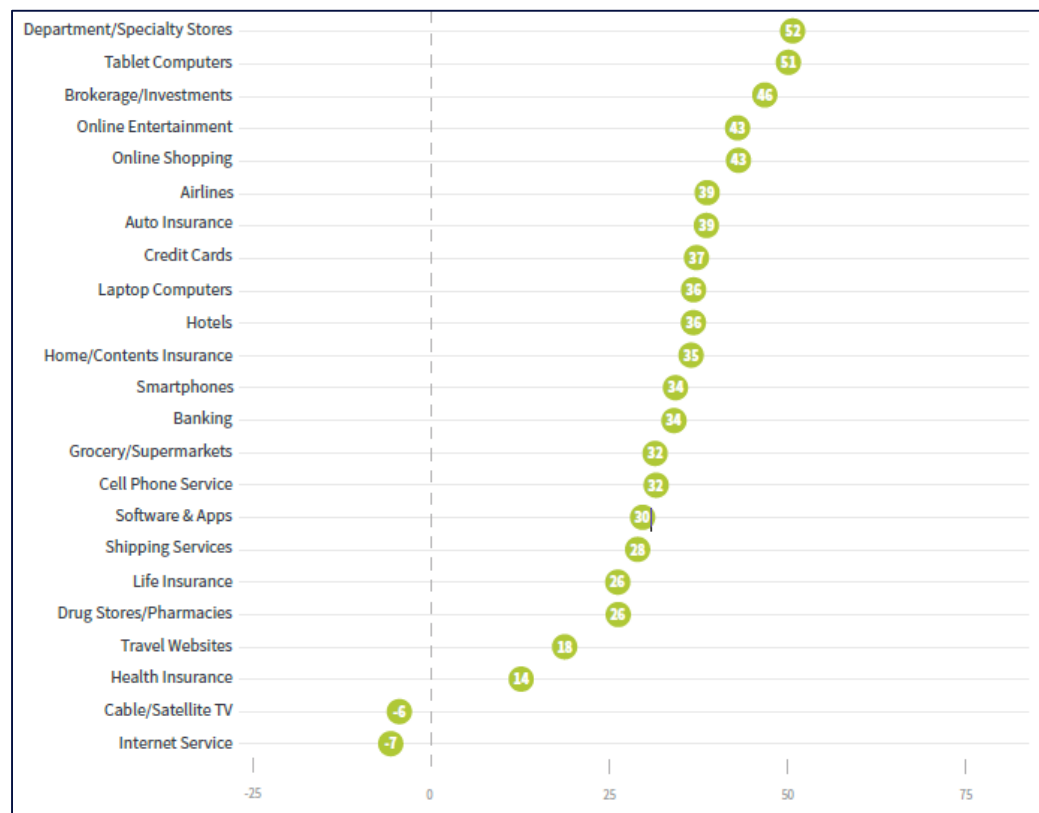
Calculating NPS score is as simple as tallying up your responses and subtracting the percentage of detractors from the percentage of promoters. The score is a whole number that ranges from -100 to 100, and indicates customer happiness with our brand experience.



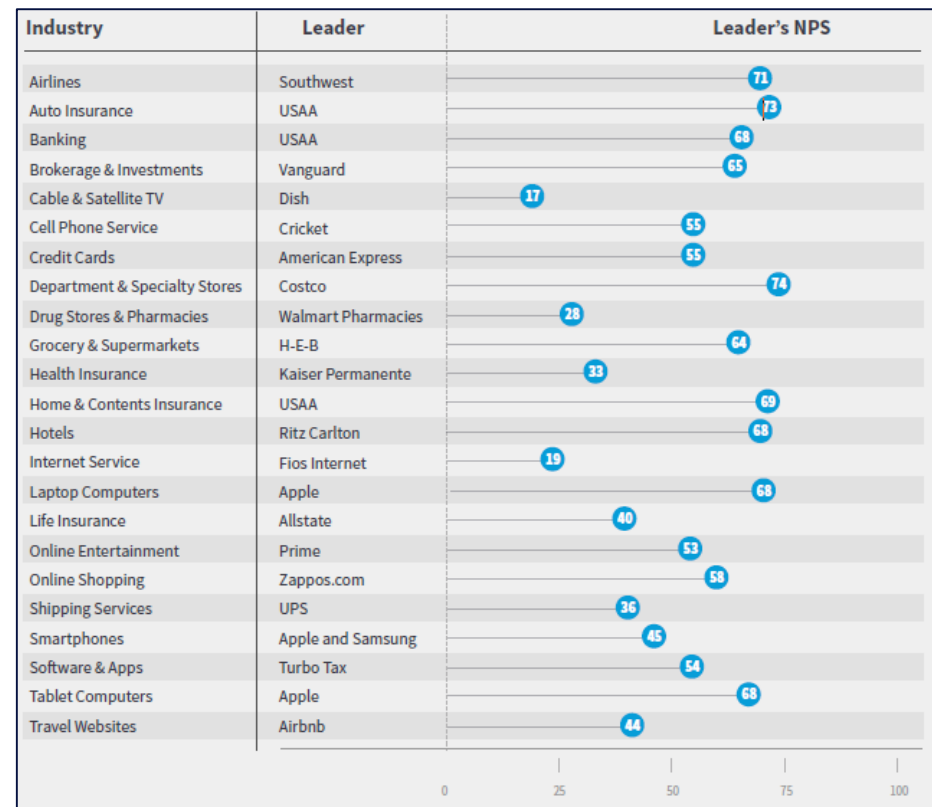
Source: How to Use Net Promoter to Drive Business Growth, Satmetrix and [NPS Calculator](#)

Benchmarks (Average NPS by Industry and Leaders)

Net Promoter Scores vary widely by industry, as you can see from the average scores for 23 industries.



Knowing what similar companies have achieved helps us to set realistic goals for improvement, and realism is key to the long-term success of your program.



Source: U.S. Consumer 2019 Net Promoter Benchmarks, Satmetrix



Importance to Business

Creating an ultimate experience that appeals to both the heart and the head is our goal.
Customers give their money, fans give their hearts of consumers.

- 91%

of marketing leaders believe that in two years companies will be competing primarily on the basis of the customer experience (Gartner)
- 44%

of consumers say that majority of customer experiences are bland
- 2%

increase in customer retention has the same effect as decreasing costs by 10%
- 69%

of consumers say that emotions count for over half their experiences
- 5X

Acquiring new customers can cost as much as 5X more than satisfying and retaining current customers

Source: U.S. Consumer 2019 Net Promoter Benchmarks, Satmetrix



Exercise 1

Identify a data science opportunity in your business context and document.

Take away

Now, I am able to

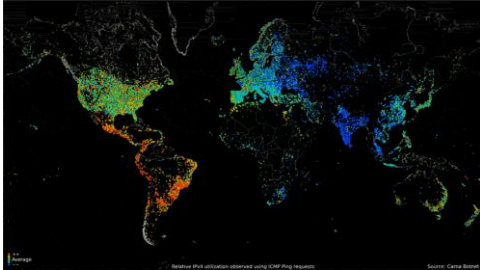
- ✓ setup the Data Science environment on IBM Cloud
- ✓ re-use the roadmap to build a machine learning system
- ✓ use a framework to assess the situation, define business objectives, and create a baseline

Feature Inputs

Time Based



Geography



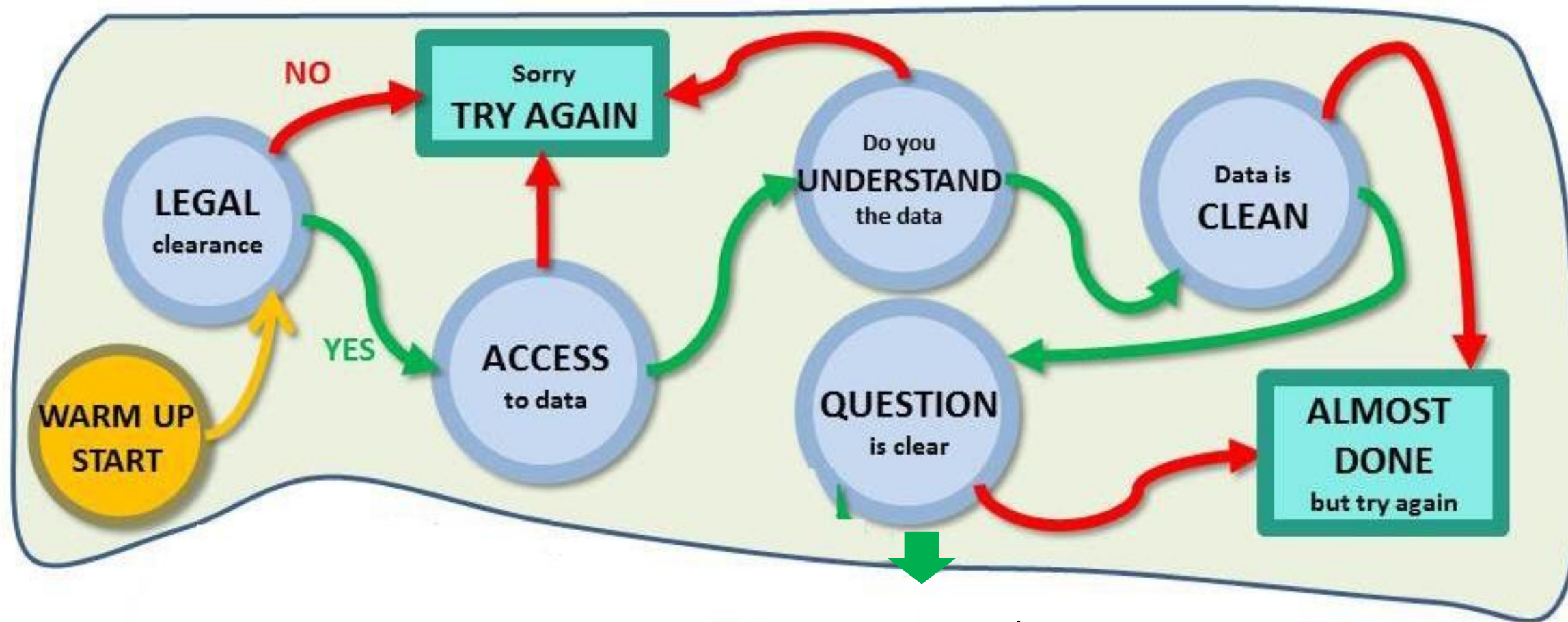
Money



Sentiment/ Emotions



Easy-Peasy



Data Preparation

Exercise 2

What data set would you gather to work on the problem statement?

Lab 1

Run the following section in the notebook.

1. Load packages, libraries and verify the version
2. Explore the data and perform quality audit

Take away

Now, I am able to

- ✓ setup the Data Science environment on IBM Cloud
- ✓ re-use the roadmap to build a machine learning system
- ✓ use a framework to assess the situation, define business objectives, and create a baseline
- ✓ load the packages, libraries, and verify the version.
- ✓ explore the data set and perform quality audit

Extract, Scale, and Reduce Dimensions

Feature Extraction

One Hot Encoding

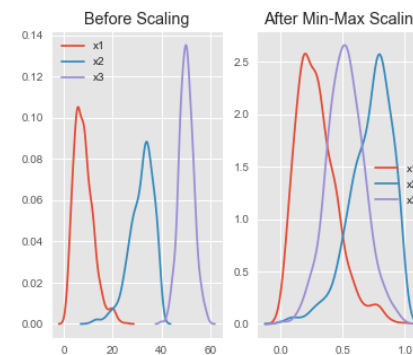


Hashing Encoder



Feature Scaling

**MinMax
Scaler**



Dimensionality Reduction

**Percent
Missing
Value**

**Amount of
Variation**

Exercise 3

How would you prepare the dataset and what challenges do you foresee?

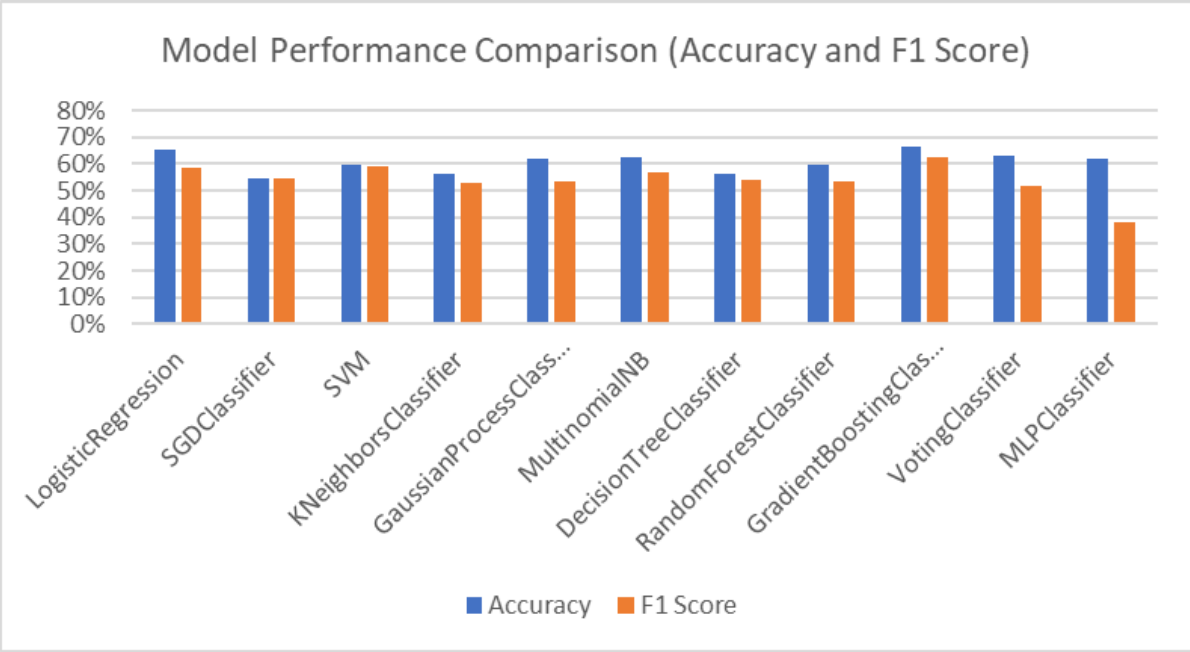
Take away

Now, I am able to

- ✓ setup the Data Science environment on IBM Cloud
- ✓ re-use the roadmap to build a machine learning system
- ✓ use a framework to assess the situation, define business objectives, and create a baseline
- ✓ load the packages, libraries, and verify the version.
- ✓ explore the data set and perform quality audit
- ✓ prepare the data for modelling, extract the features, scale the data, and reduce the dimensions

Model Selection and Evaluation

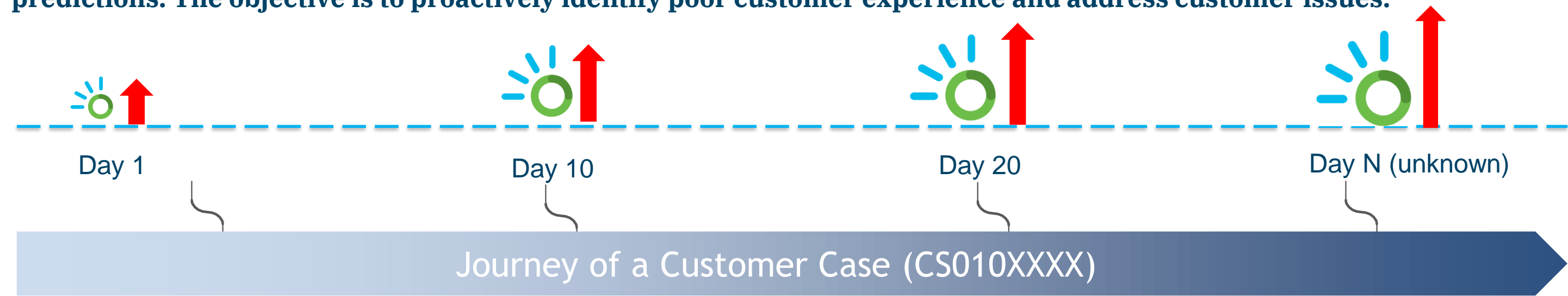
Model Name	Metrics				
	Accuracy	Precision	Recall	F1 Score	
LogisticRegression	📈 65%	63%	59%	📈	58%
SGDClassifier	📉 55%	59%	58%	📉	55%
SVM	📈 60%	59%	60%	📈	59%
KNeighborsClassifier	📈 56%	53%	53%	📉	53%
GaussianProcessClassifier	📈 62%	57%	55%	📉	53%
MultinomialNB	📈 62%	59%	57%	📈	57%
DecisionTreeClassifier	📈 56%	54%	54%	📉	54%
RandomForestClassifier	📈 60%	55%	54%	📉	53%
GradientBoostingClassifier	📈 67%	64%	62%	📈	62%
VotingClassifier	📈 63%	59%	55%	📉	52%
MLPClassifier	📈 62%	31%	50%	📉	38%



Demonstration



The algorithm consumes multiple signals (time, geography, spend, and sentiments) and gives the non promoter predictions. The objective is to proactively identify poor customer experience and address customer issues.



Lab 2

Run the following section in the notebook.

7. Split data into train and test sets

8. Model Selection

9. Performance Metric

10. Evaluation

Exercise 4 & 5

What modeling techniques would you attempt and metrics would you use to evaluate the model performance?

Take away

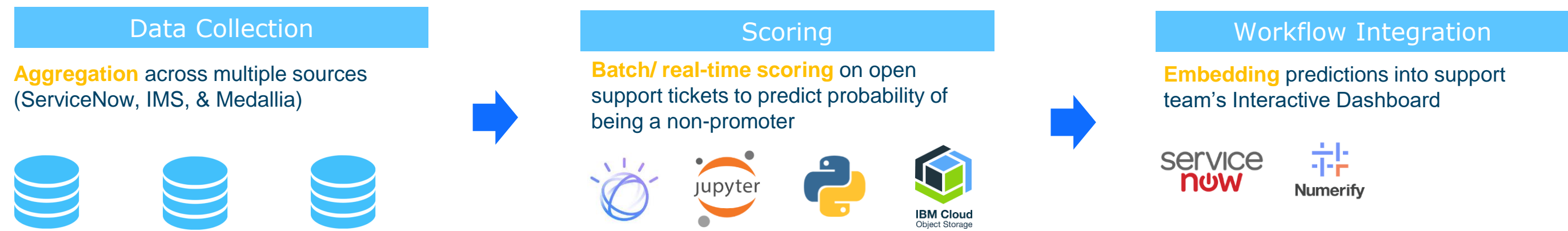
Now, I am able to

- ✓ setup the Data Science environment on IBM Cloud
- ✓ re-use the roadmap to build a machine learning system
- ✓ use a framework to assess the situation, define business objectives, and create a baseline
- ✓ load the packages, libraries, and verify the version.
- ✓ explore the data set and perform quality audit
- ✓ prepare the data for modelling, extract the features, scale the data, and reduce the dimensions
- ✓ Split data into train & test sets, select model, and evaluate performance metrics

Proposed Solution

The model developed as a part of the hack uses artificial intelligence and machine learning to predict Non Promoters on historical data pattern. Key aspects of approach includes,

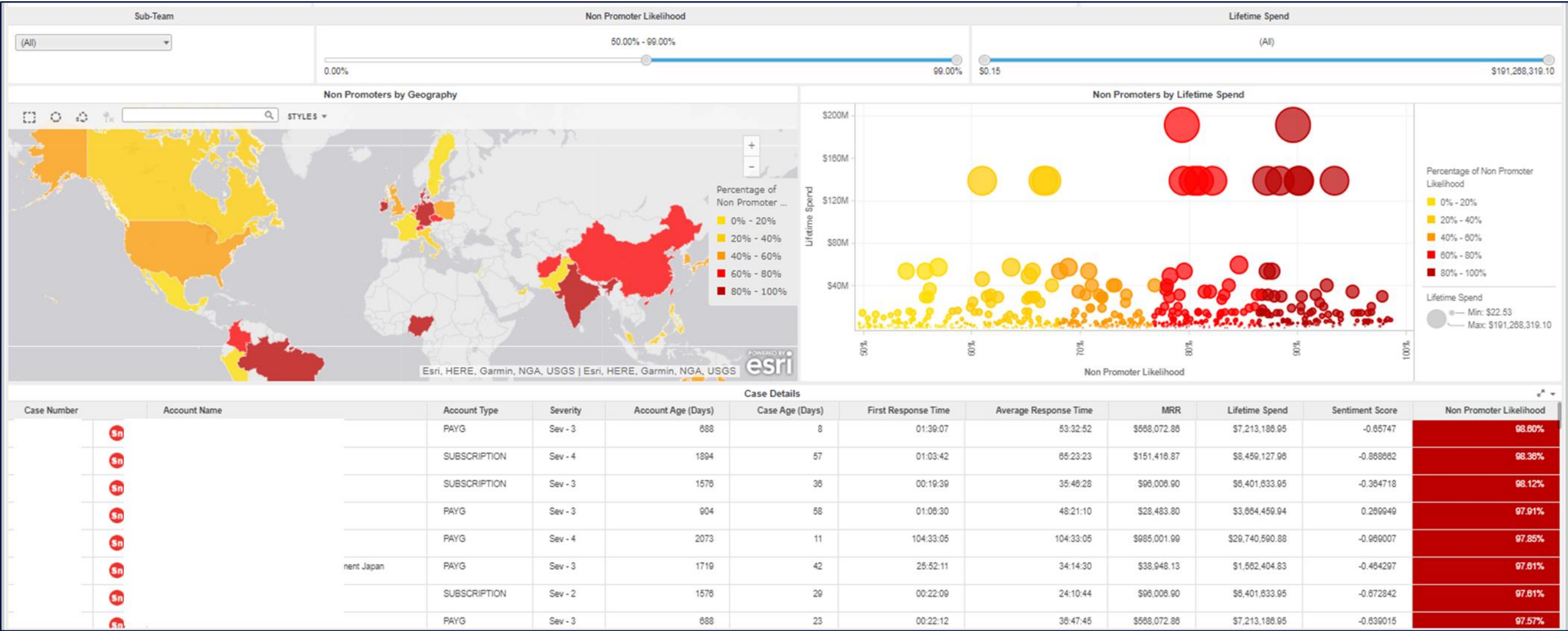
- 1. Watson NLP to create additional features from customer conversation logs
- 2. Machine Learning algorithms to come up with the predictions



Solution Integration

“Ideas are easy, Execution is everything.” John Doerr

NPS Predictions Board



Note: This visual highlighted above is crafted for workshop purpose.

Lab 3

Run the following section in the notebook.

11. Deployment

Exercise 6

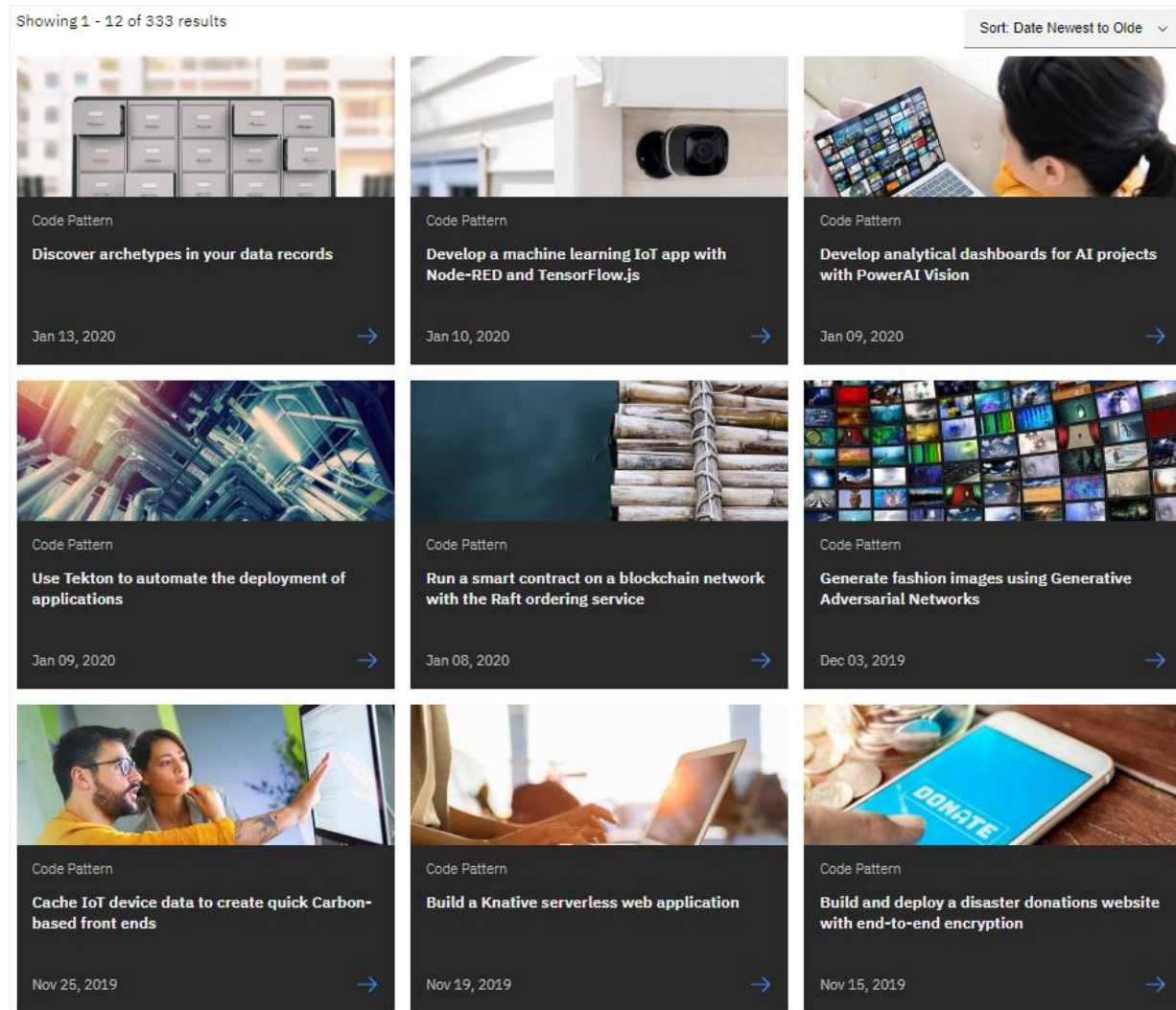
How do you plan to consume the outputs of model?

Take away

Now, I am able to

- ✓ setup the Data Science environment on IBM Cloud
- ✓ re-use the roadmap to build a machine learning system
- ✓ use a framework to assess the situation, define business objectives, and create a baseline
- ✓ load the packages, libraries, and verify the version.
- ✓ explore the data set and perform quality audit
- ✓ prepare the data for modelling, extract the features, scale the data, and reduce the dimensions
- ✓ Split data into train & test sets, select model, and evaluate performance metrics
- ✓ generate ideas on how to consume the predictions and integrate solution in business systems.

IBM Code Patterns



Link: <https://developer.ibm.com/patterns/>

Contact Us

	<p>Name: Maureen Norton Email Id: nortonm@us.ibm.com LinkedIn: https://www.linkedin.com/in/maureennorton/</p>
	<p>Name: George Stark Email Id: gstark@us.ibm.com LinkedIn: https://www.linkedin.com/in/gestark/</p>
	<p>Name: Neeraj Madan Email Id: nmadan@us.ibm.com LinkedIn: https://www.linkedin.com/in/neerajmadan/</p>

