IBM Watson Studio

Kavya Ahuja
Lakshmi Yetukuri
Hannah Adamson

About IBM Watson

- IBM’S software platform for data science
- Lets users collaborate and access analytical models
- Uses languages like R, Python, Scala
- Brings together open source programs like Rstudio and Python into a integrated environment

Set Up Instructions

Cloud vs. IBM Studio Desktop

- With Cloud you can deploy to more environments like JupyterLab, Rstudio, Git, AWS,etc
- With Cloud, you can collaborate
- Desktop -includes Jupyter Notebook but not Lab

Cloud Instructions:

- Click on https://cloud.ibm.com/catalog/services/watson-studio
- Create account (Lite account is free)

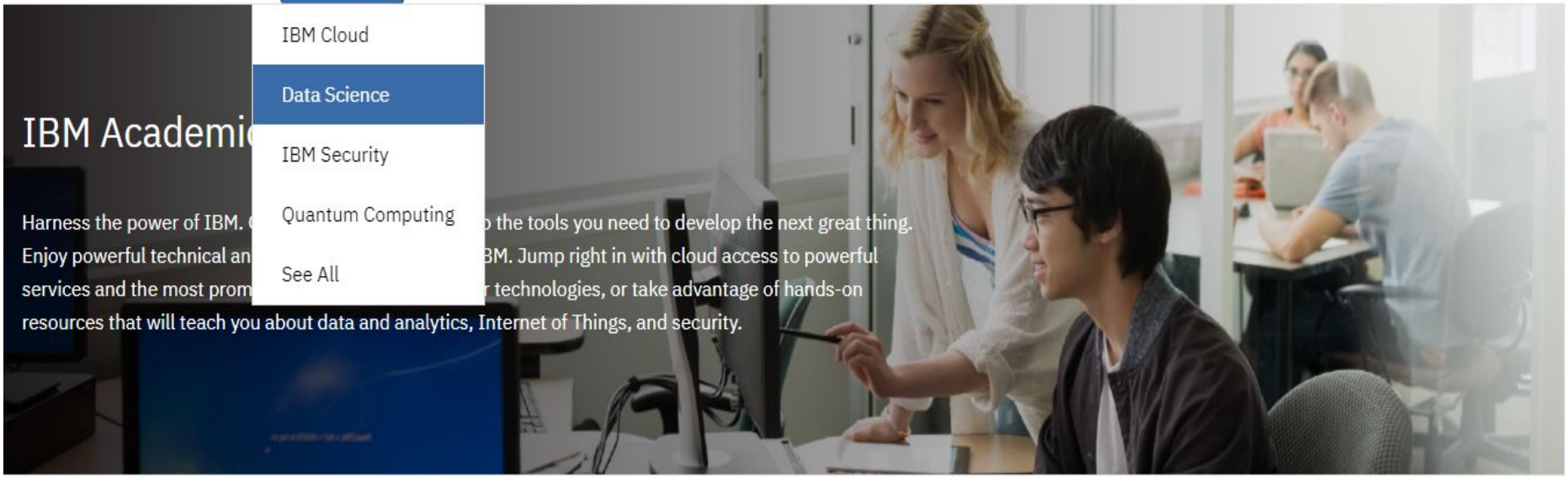
Desktop Instructions:

- Free for students
- Create a account using ibm.biz/academic

IBM

IBM Academic Initiative | Technology | Discipline | Usage terms | Additional Resources | Support

- IBM Cloud
- Data Science
- IBM Security
- Quantum Computing
- See All



Use your skills to take on climate change and join a community of developers and innovators using open source technology in the 2021 #CallforCode challenge.

What is the IBM Academic Initiative?

Watson Studio Desktop

Integrated SPSS Modeler functionality to drag-and-drop your way to ML, model building, and data exploration. Data science students can use Watson Studio Desktop to exercise their R and Python skills with Jupyter Notebooks. Use the integrated SPSS Modeler and Data Shaper to explore and chart data.

Download →

Product Information →

Stack Overflow →

ILOG CPLEX Optimization Studio

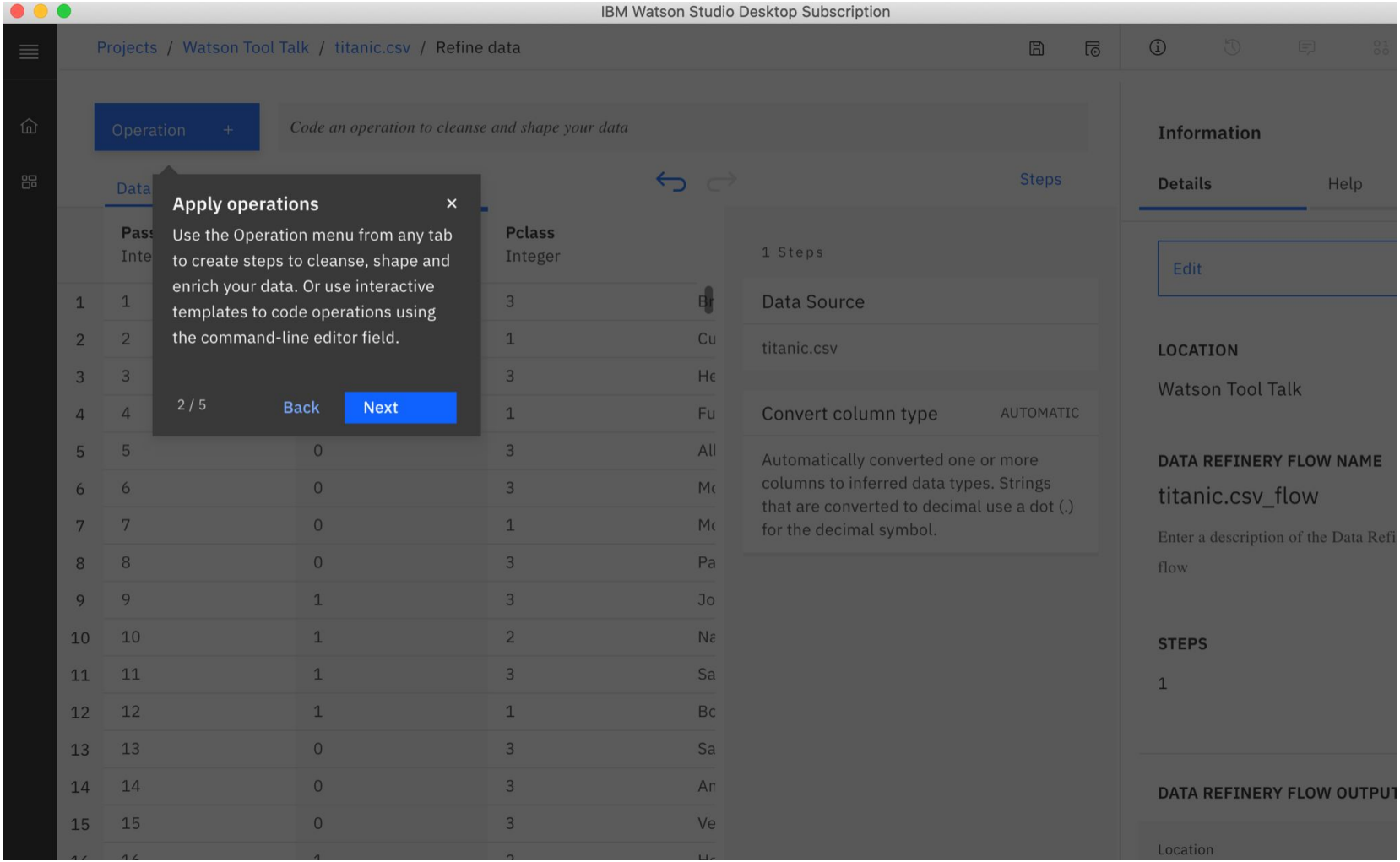
Analytical decision support toolkit for rapid development and deployment of optimization models using mathematical and constraint programming. It combines an integrated development environment with the powerful Optimization Programming Language and high-performance CPLEX and CP Optimizer solvers.

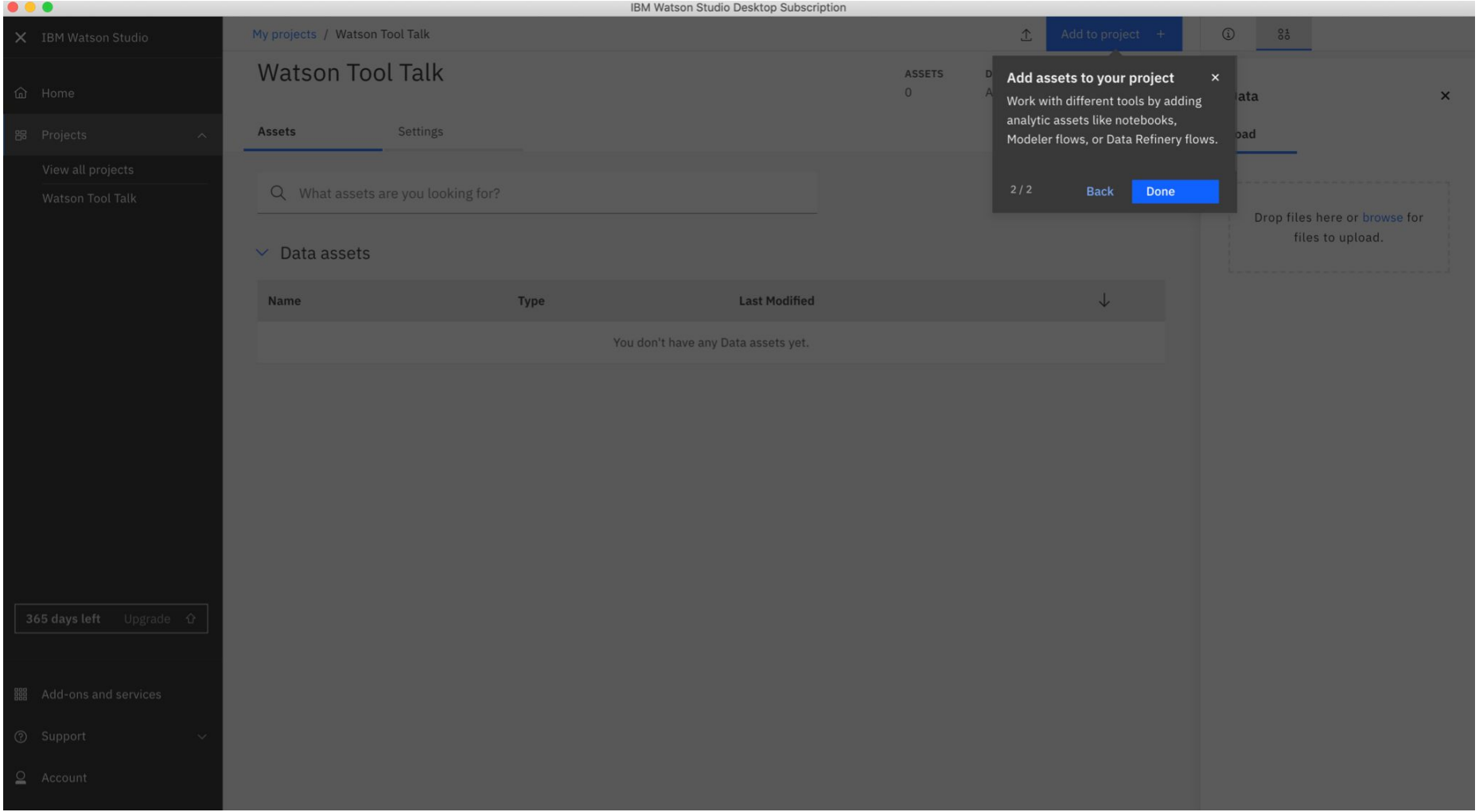
Download →

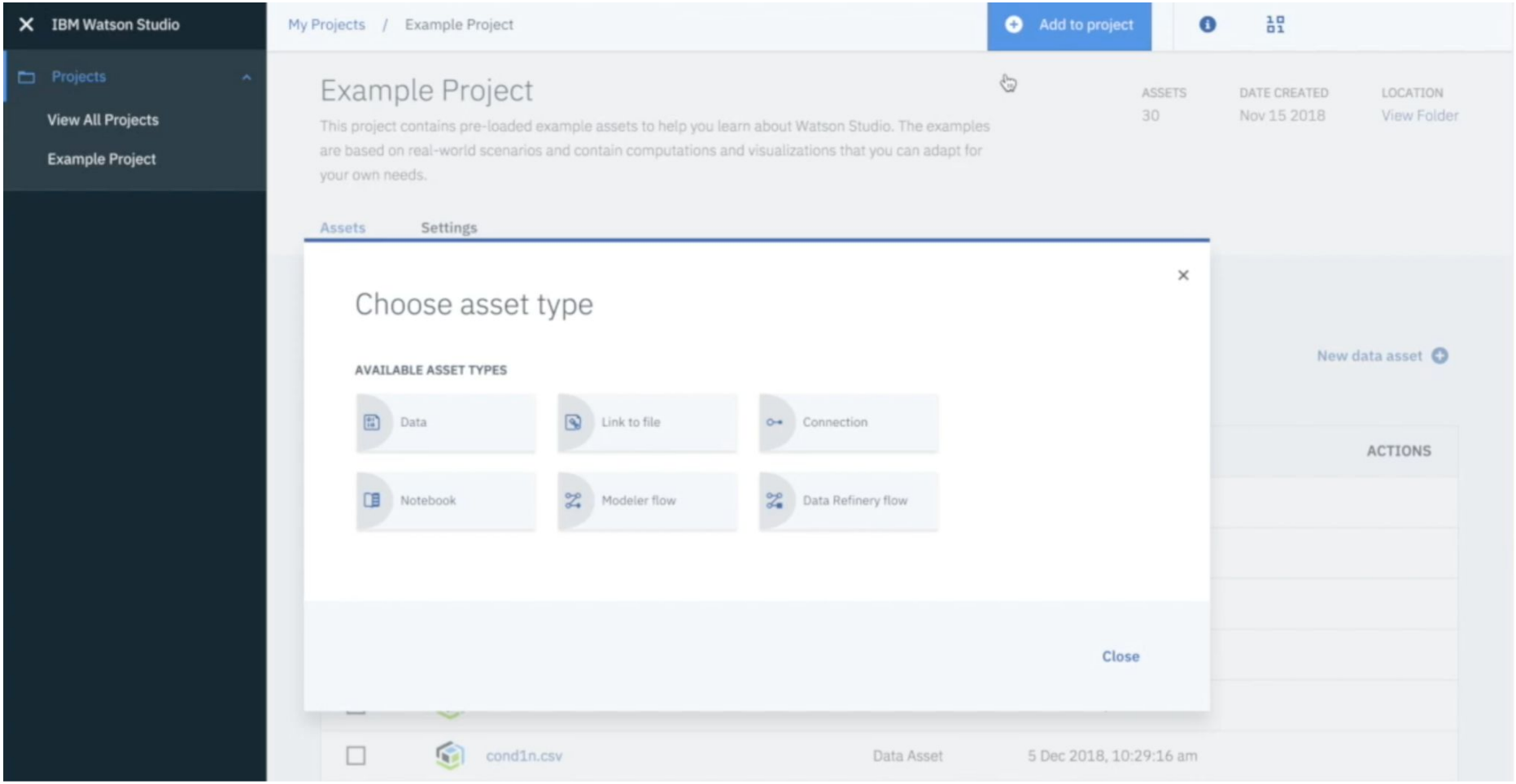
Product Information →

CPLEX Community →

Data Preparation on Watson


IBM Watson Studio Desktop Subscription
Projects / Watson Tool Talk / titanic.csv / Refine data
Operation +
Code an operation to cleanse and shape your data
Steps
Apply operations
Use the Operation menu from any tab to create steps to cleanse, shape and enrich your data. Or use interactive templates to code operations using the command-line editor field.
2 / 5
Back
Next
Pclass
Integer
1 Steps
Data Source
titanic.csv
Convert column type
AUTOMATIC
Automatically converted one or more columns to inferred data types. Strings that are converted to decimal use a dot (.) for the decimal symbol.
Information
Details
Help
Edit
LOCATION
Watson Tool Talk
DATA REFINERY FLOW NAME
titanic.csv_flow
STEPS
1
Location


IBM Watson Studio Desktop Subscription
IBM Watson Studio
Home
Projects
View all projects
Watson Tool Talk
365 days left
Upgrade
Add-ons and services
Support
Account
My projects / Watson Tool Talk
Add to project
Watson Tool Talk
ASSETS
0
Assets
Settings
What assets are you looking for?
Data assets
Name
Type
Last Modified
You don't have any Data assets yet.
Add assets to your project
Work with different tools by adding analytic assets like notebooks, Modeler flows, or Data Refinery flows.
2 / 2
Back
Done
Drop files here or browse for files to upload.


IBM Watson Studio
Projects
View All Projects
Example Project
My Projects / Example Project
Add to project
Example Project
This project contains pre-loaded example assets to help you learn about Watson Studio. The examples are based on real-world scenarios and contain computations and visualizations that you can adapt for your own needs.
ASSETS
30
DATE CREATED
Nov 15 2018
LOCATION
View Folder
Assets
Settings
Choose asset type
AVAILABLE ASSET TYPES
Data
Link to file
Connection
Notebook
Modeler flow
Data Refinery flow
Close
New data asset
ACTIONS
cond1n.csv
Data Asset
5 Dec 2018, 10:29:16 am

IBM Watson Studio

Projects

View All Projects

Example Project

Support

My Projects / Example Project / 017-SparkInsightsintotheChicago...

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3

Run Format Markdown

Chicago Car Accident Data Analysis

[Notebook from: http://bit.ly/byte-size-data-science]
In this notebook, we analyze the data using a Python environment.
We also use Pixiedust as the engine over Mapbox to display maps in the later part of the analysis.

In an additional section, we see how we could use additional data to add the city name to each record.

Additional Information

The chicago accident information includes three files: Crashes, people, and vehicles.

In this notebook, we explore the crashes through a file called `ChicagoTrafficCrashes20180917.csv`

Read the crash data

In this section, we read the data as a Spark DataFrame

In [ ]:
```
# PixieDust is an open source library that was contributed by IBM
!pip install --user --upgrade pixiedust
```

In [ ]:
```
import pixiedust
```

In [ ]:
```
from pyspark.sql import SparkSession
import urllib.request
import zipfile

spark = SparkSession.builder.getOrCreate()
```

In [ ]:
```
url = 'https://github.com/jacquesroy/byte-size-data-science/raw/master/data/ChicagoTrafficCrashes20180917.csv.zip'
# get the filename from the url: "ChicagoTrafficCrashes20180917.csv"
zipfilename = url.rsplit('/', 1)[-1]
filename = zipfilename.rsplit('.', 1)[0]
```

Jupyter Notebook with Cloud Version

Why use IBM Watson

- The simplest way to visualize data set
- Many tools to modify data set
- Significantly decreases runtime for long data sets
- Can use open source software like Jupyterlab and Anaconda without downloading the software
- Can access your work with any computer
- Can use collaborators