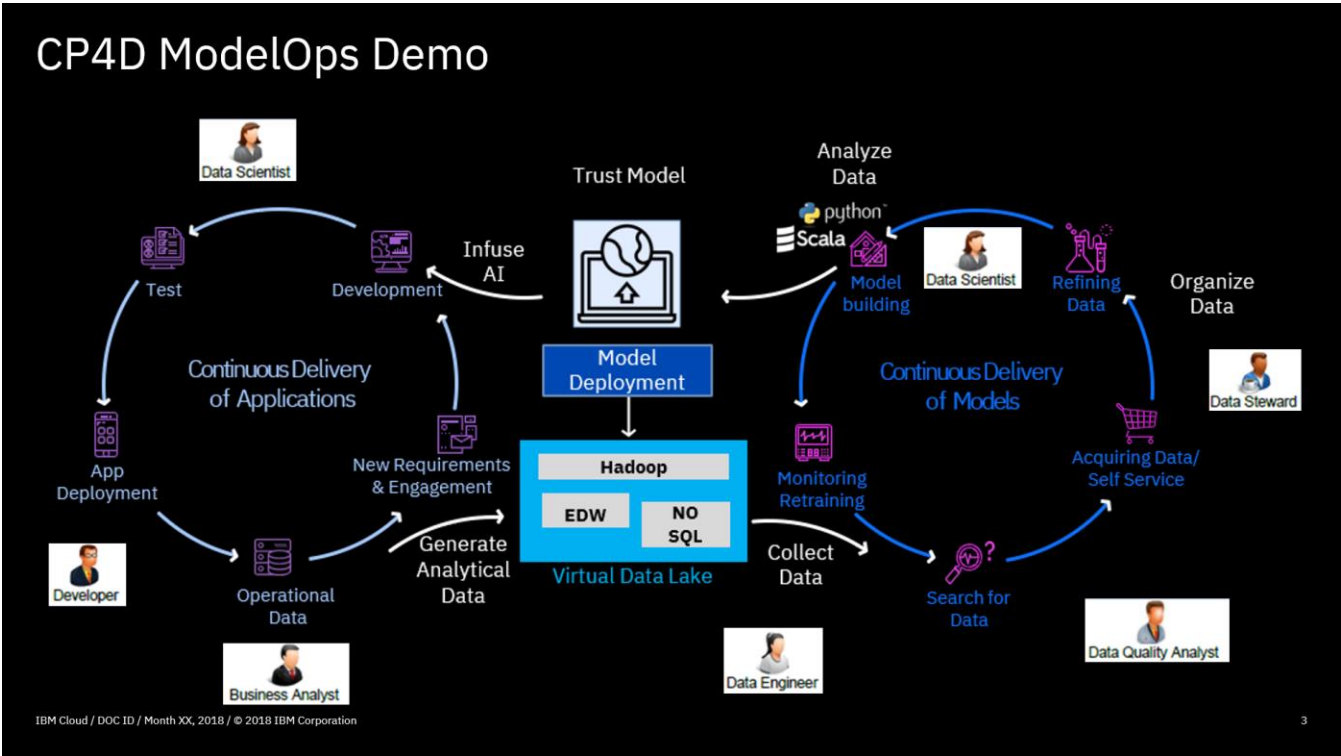


CP4D Industry Accelerator - Reducing Severe Weather Risk For Insurance

Using The Cloud Pak For Data Model Ops Platform To Reduce Severe Weather Risk	1
CP4D Analytic Artifacts Included in This Accelerator	2
How To Import The Industry Accelerator?.....	12
How To Import The Watson Studio Project, Dashboard and Modeler Flow	14
Import The Project	14
Import the Analytics Dashboard	16
Importing the SPSS Modeler Stream and CP4D Modeler Flow.....	20
Using AutoAI For Model Selection And Analysts That Want Data Science But Aren't Data Scientists.....	24

Using The Cloud Pak For Data Model Ops Platform To Reduce Severe Weather Risk

Severe weather has a significant impact on insurers and their policyholders. This industry accelerator simulates a point in history where there were billions of dollars of damage caused by fire, wind and hail 2” in size within the south western part of the United States. This accelerator will show how data engineers, data stewards, data quality analysts, data scientists, citizen data scientists, analysts, and developers can all focus on their role to reduce severe weather risk and work together at the same time using Cloud Pak For Data as the ModelOps Platform.



- Collect, ingest and merge historical severe weather (hail, wind, and fire) data, coming The Weather Company Data Packages, with historical policyholder claim data to understand what the weather was in the past when a claimed occurred. The same data will be used downstream to build test and deploy claim damage prediction models and data built by data scientists and analysts.
- Organize your analytic artifacts such as data, notebooks, models, data source connections and dashboards making it easy for others to find and collaborate through governed trusted catalogs
- Analyze as a data scientist or analyst, to build, test and deploy a model to predict claim damage or develop analytic dashboards to show many views of severe claims damage.
- Infuse model results into downstream dashboards, online/batch scoring and other applications such as ESRI maps.

CP4D Analytic Artifacts Included in This Accelerator

The **Analytic Dashboard** allows insurers to understand the historic and predicted impact of severe weather on their policyholders by helping them know how much cash they should have on hand to survive a severe event, how to prepare for an upcoming event and know which customers have been or will be impacted the most. In this case, Colorado was hit the worst, so that's where we need to focus our efforts first.

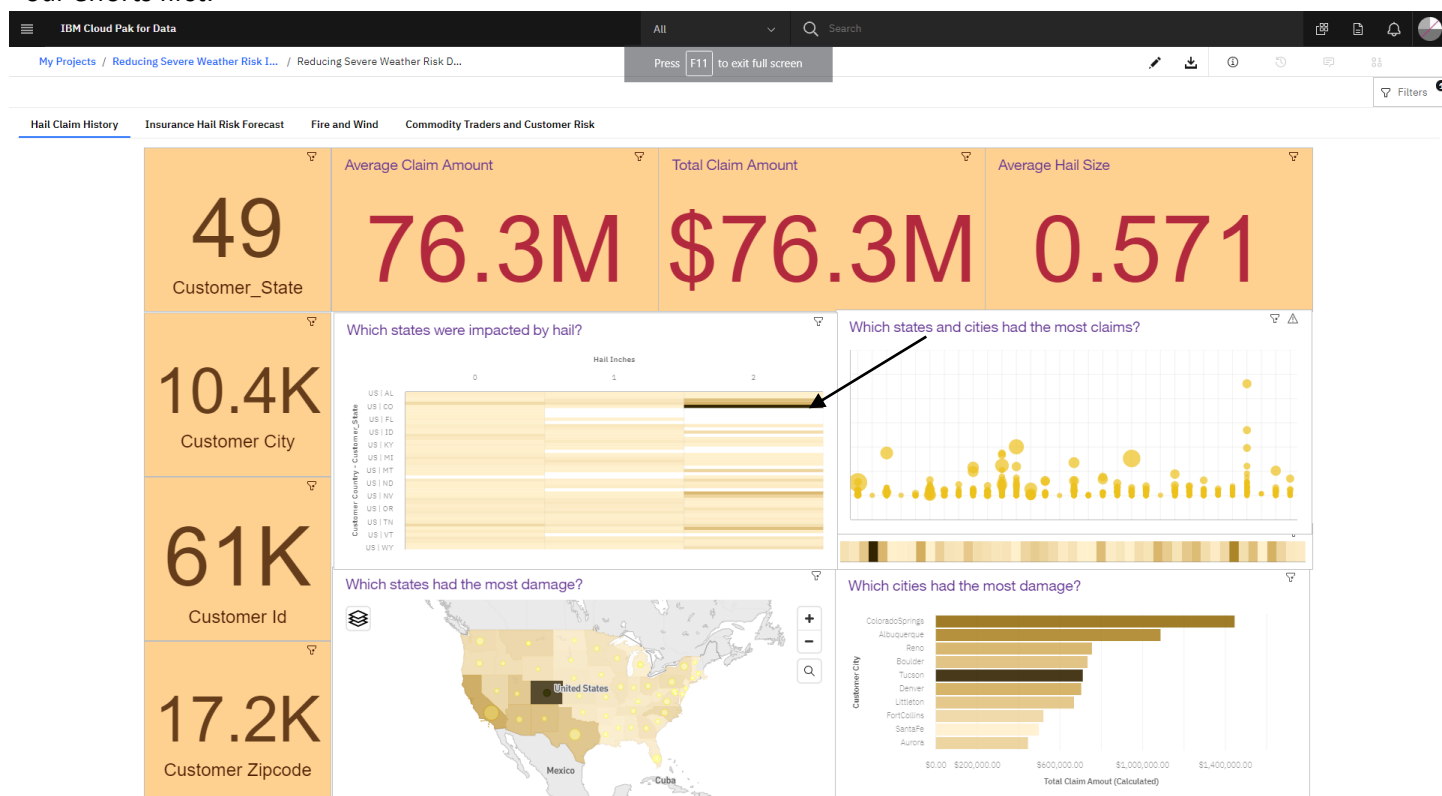


Figure 1- Cloud Pak For Data Analytics Dashboard – Tab 1 - Hail Claim History Highlighting most damage occurred with 2" of hail in CO.

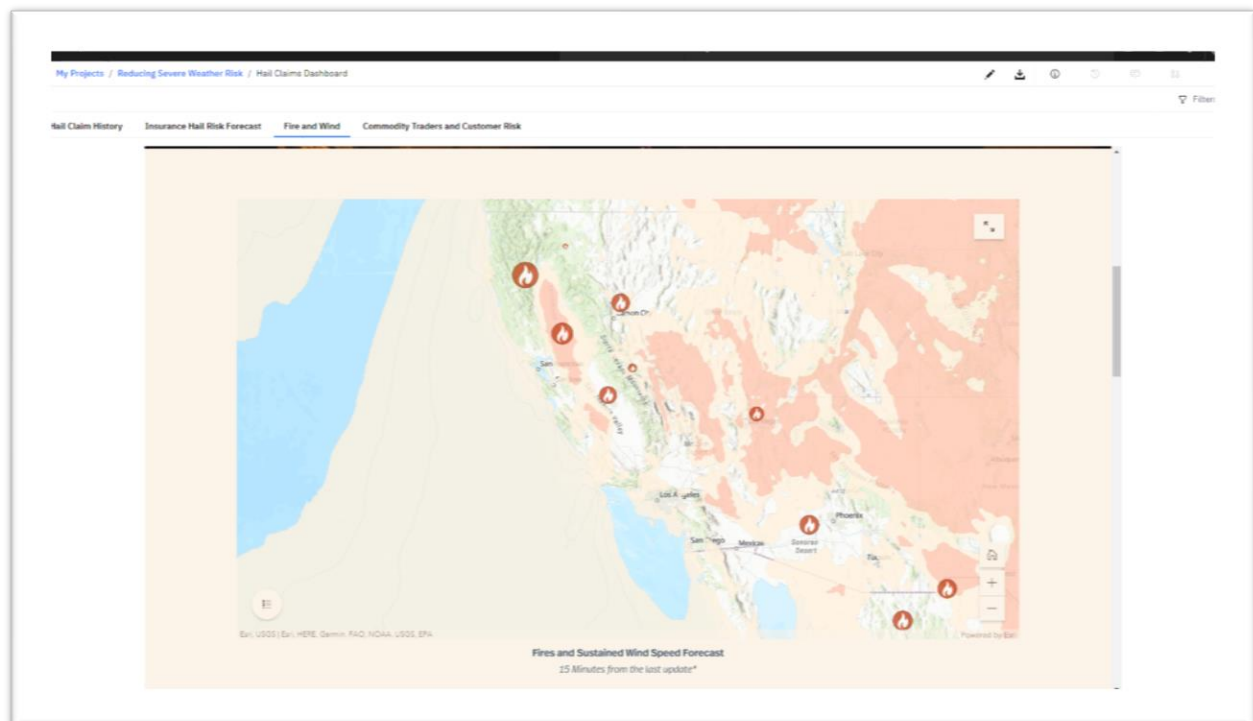
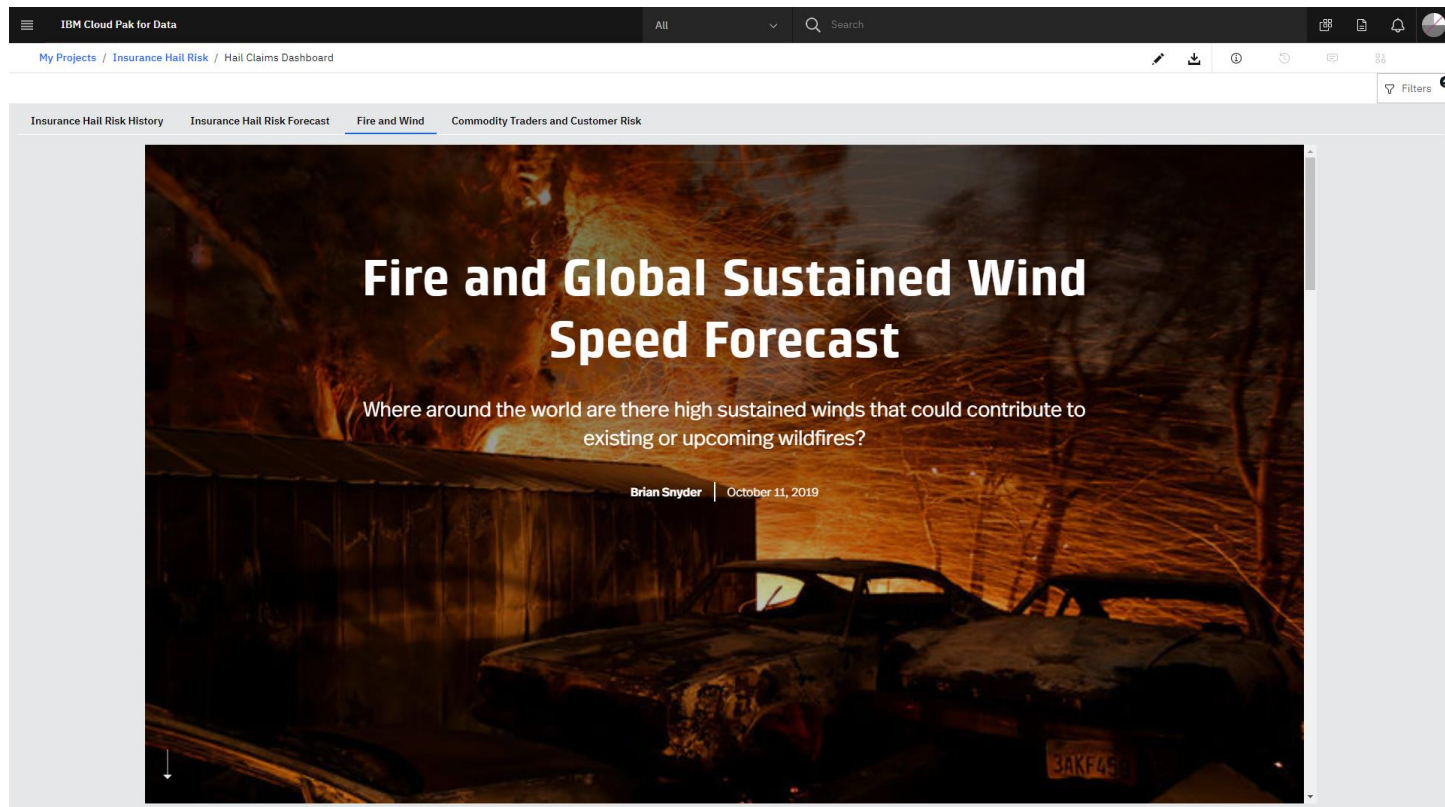


Figure 2 - Cloud Pak For Data Analytics Dashboard - Fire and Severe Wind in the SW United States forecasted for the next 15, 30 and 45 minutes. CP4D was used to pull weather polygons and other geospatial data from The Weather Company and update ESRI maps using CP4D Python Notebooks on a regular schedule to provide near real time info for those concerned with severe weather. For more information on using weather data @ IBM go here: <https://www.ibm.com/weather>

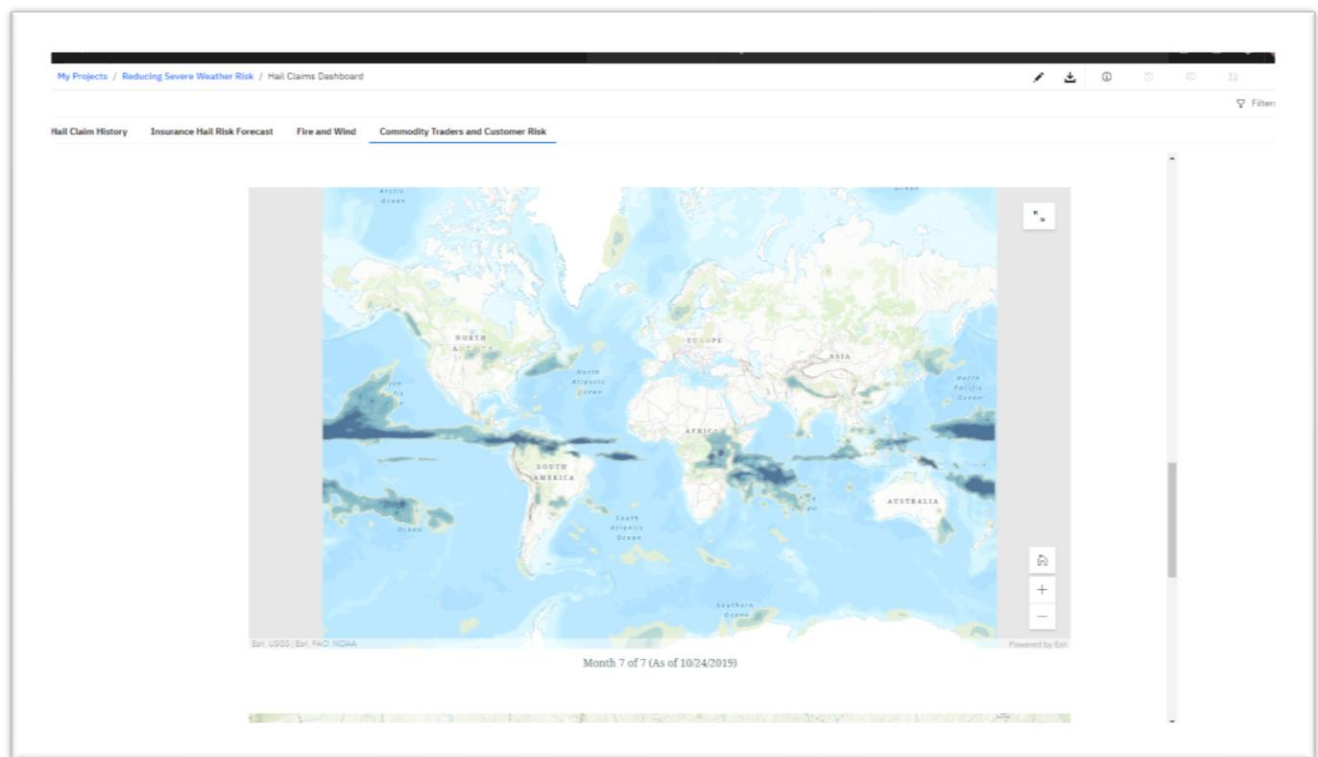
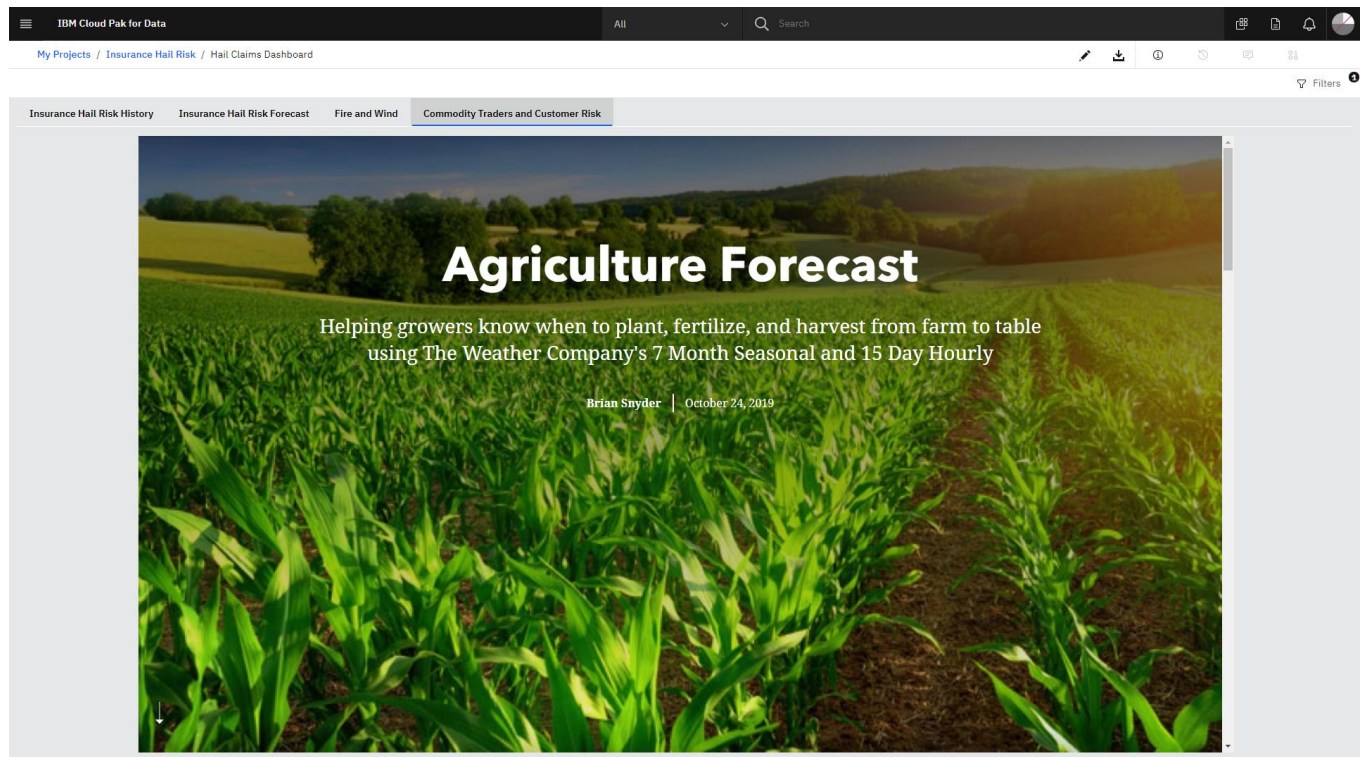


Figure 3 - Cloud Pak For Data Analytic Dashboard - 7 Month Seasonal Forecast for traders and those interested in investment risk for longer and short term investment strategies. A scheduled Python Notebook updates this map on a regular basis pulling weather polygons and other geospatial data from The Weather Company Data Packages and pushing them to ESRI maps using the ESRI API. For more information on using weather data @ IBM go here: <https://www.ibm.com/weather>

IBM Cloud Pak for Data

All

Search

Catalogs / My Catalog

Add to Catalog

hail

Any type

Any source

Any tag

Clear all

Watson Recommends

Highly Rated

Recently Added

Collapse

Dashboard

Credit Card Fraud Dash...

Owner: Brian Snyder

Added: Sep 02, 2020 6:47 PM

Tags: fraud cred...

☆☆☆☆☆ 0 reviews

Notebook

Credit Card Fraud Detection

Owner: Brian Snyder

Added: Sep 02, 2020 6:47 PM

Tags: fraud cred...

☆☆☆☆☆ 0 reviews

Data asset

suggestedFraudCharges.csv

Owner: Brian Snyder

Added: Sep 02, 2020 6:47 PM

Tags: fraud trans...

☆☆☆☆☆ 0 reviews

Data asset

december transactions.csv

Owner: Brian Snyder

Added: Sep 02, 2020 6:46 PM

Tags: fraud trans...

☆☆☆☆☆ 0 reviews

Data asset

december transactions_wit...

Owner: Brian Snyder

Added: Sep 02, 2020 6:46 PM

Tags: fraud trans...

☆☆☆☆☆ 0 reviews

Notebook

TradingCustomerChurnCla...

Owner: Brian Snyder

Added: Aug 31, 2020 7:56 AM

Tags: python churn

☆☆☆☆☆ 0 reviews

Dashboard

Mon

Owner: Added

Tags:

☆☆☆☆☆ 0 reviews

Showing 8 of 8 items

<input type="checkbox"/>	Name	Owner	Tags	Business Terms	Type	Date Added
<input type="checkbox"/>	ClaimsLocationsHail.csv	BS Brian Snyder	claims insurance + 1 more		Data asset	Aug 04, 2020
<input type="checkbox"/>	ClaimsLocationsHail.csv	BS Brian Snyder	claims locations + 1 more		Data asset	Aug 04, 2020
<input type="checkbox"/>	CustomerReportedHail.csv	BS Brian Snyder	hail historical weather + 2 more		Data asset	Aug 04, 2020
<input type="checkbox"/>	HAIL_CUSTOMER_HAIL_AMOUNTS	BS Brian Snyder	connected-data		Data asset	Aug 04, 2020
<input type="checkbox"/>	HAIL_CUSTOMER_LOCATIONS	BS Brian Snyder	connected-data		Data asset	Aug 04, 2020
<input type="checkbox"/>	HAIL_CUSTOMER_LOCATIONS_CLAIMS_AND_HAIL	BS Brian Snyder	connected-data		Data asset	Aug 04, 2020
<input type="checkbox"/>	HAIL_DATABASE	BS Brian Snyder	global_connection		Connection	Aug 04, 2020
<input type="checkbox"/>	Hail Claims Dashboard	BS Brian Snyder	insurance hail + 3 more		Dashboard	Jul 24, 2020

Figure 4 Organize Approved and Regulated Data Artifacts. In this case, make it easy to search for any analytic artifacts related to hail.

Auto AI suggests to a data scientist, a citizen data scientist, or an analyst to use the *Extra Trees Regression* model when trying to predict Damage Claim Amount from severe weather based on Root Square Mean Error that may include hyperparameterization and enhancements to improve model accuracy. It's ability to leverage IBM proprietary model selection and data preparation accelerates time to insight with explainability that significantly accelerates time to insight.

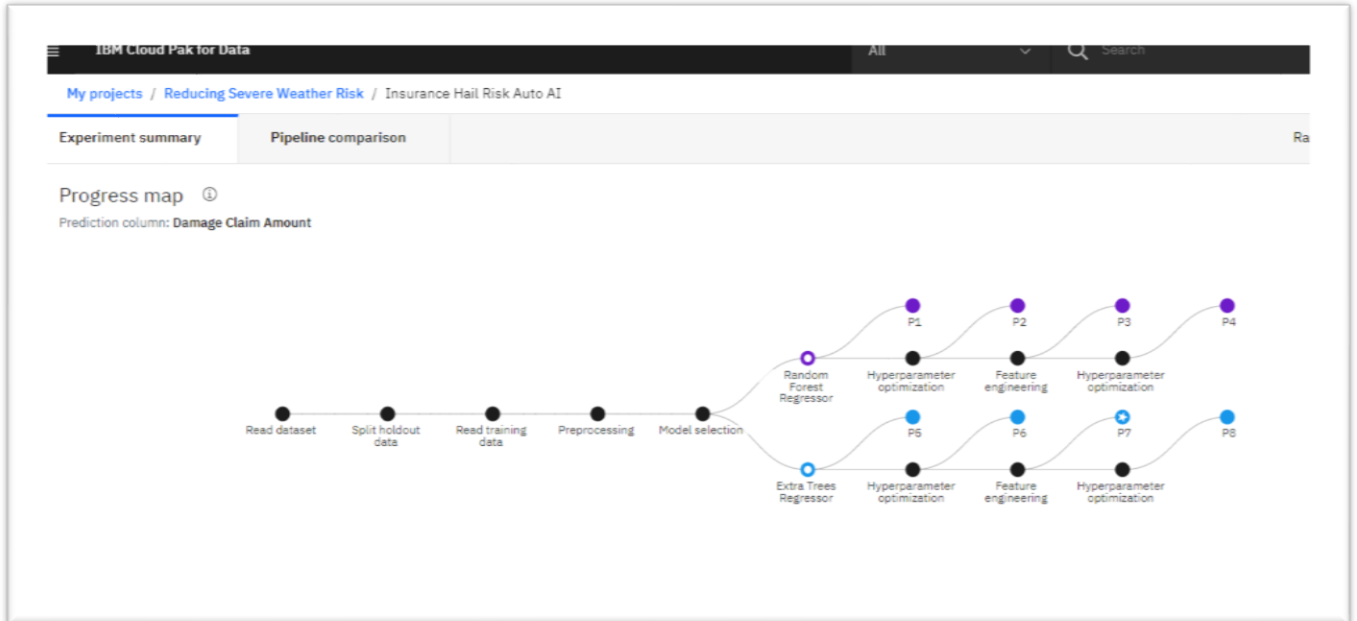


Figure 5 Automated data processing, IBM proprietary data prep and best fit model selection with explainability by providing a no blackbox way of exporting the preferred model to Python.

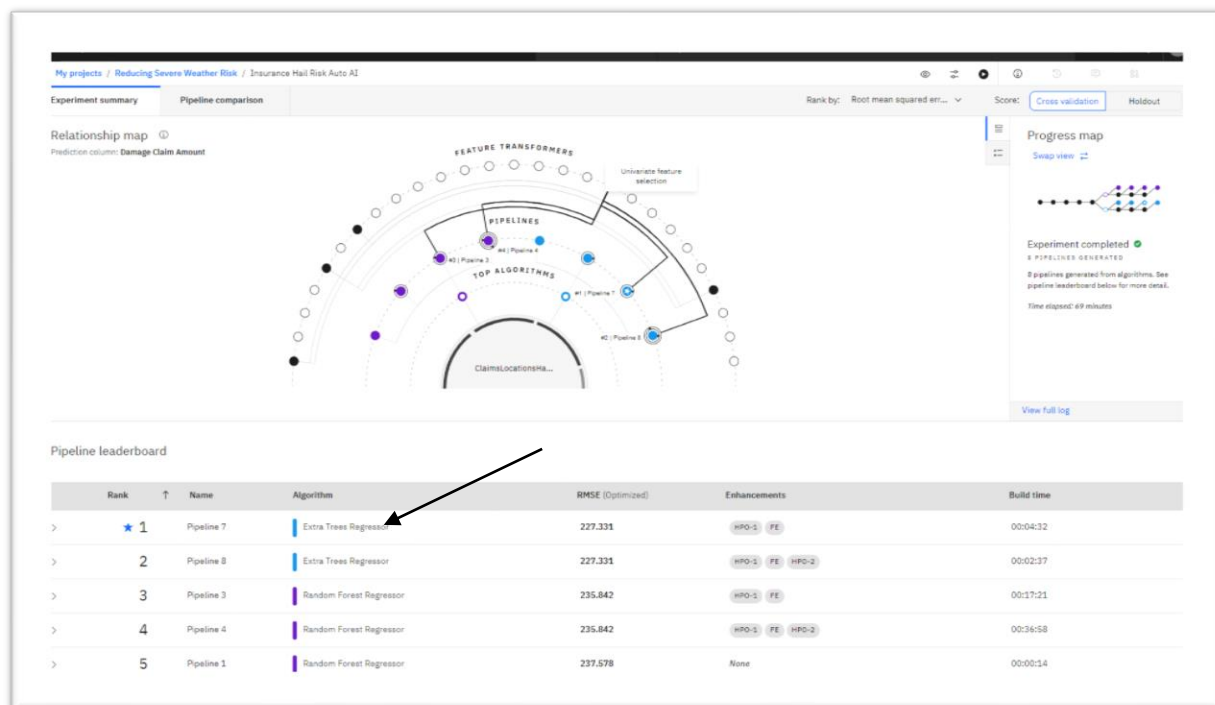


Figure 6 Pipeline shows top 10 best fit models to predict Damage Claim Amount

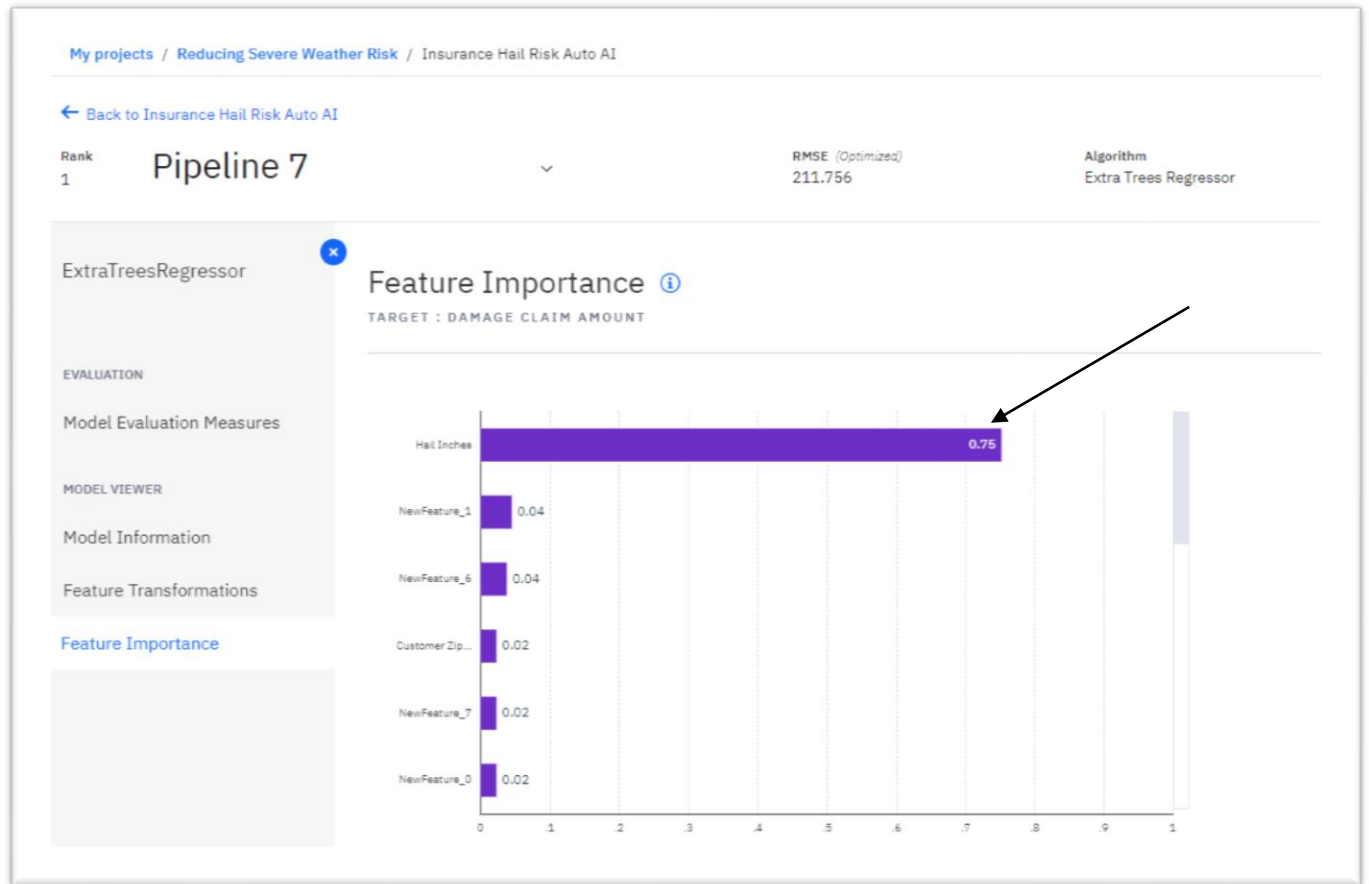


Figure 7 The best fit model selected uses the Extra Trees Regressor algorithm suggests that hail is the leading contributor to damage.

Now that they know a recommended model to use, here's examples that a data scientist and citizen data scientist will use to build their own *Extra Trees Regression* model using **IBM Watson Studio Python (Or R)**, **IBM SPSS Modeler Desktop**, and/or **CP4D Modeler Flows** showcasing the various options to perform data mining, data preparation, data transformation, modeling, and deployment of a model. IBM SPSS Modeler Desktop and CP4D Modeler Flows provide data science capabilities without the need for code. Code (Python, R) can be brought to them if needed too.

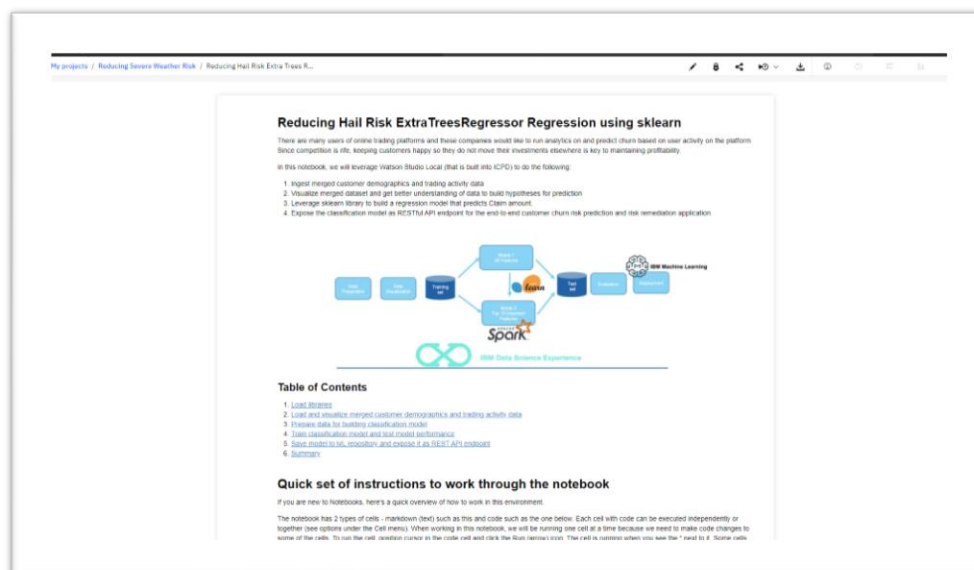


Figure 8 Jupyter Python Notebook Using Sklearn Machine Learning and WatsonMachineLearning to promote model to Watson Machine Learning. The model's purpose is to predict Claim Damage Amounts.

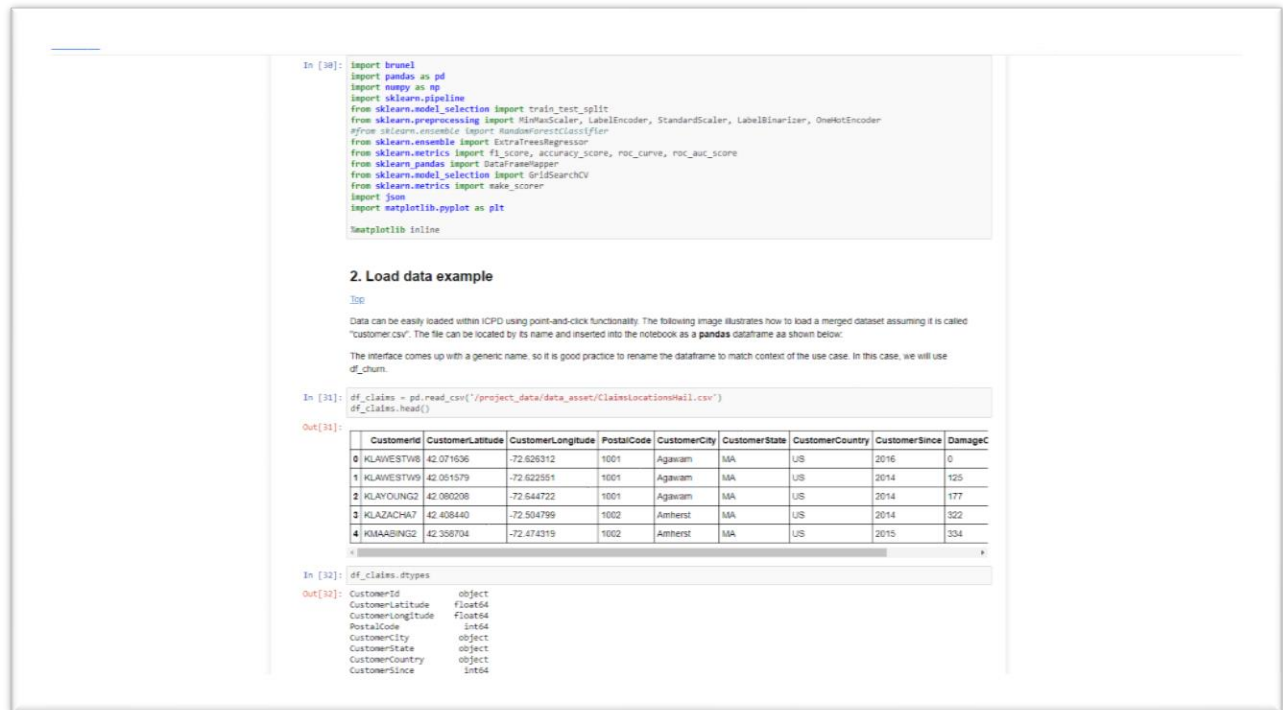


Figure 9 Loading data into Python Notebook

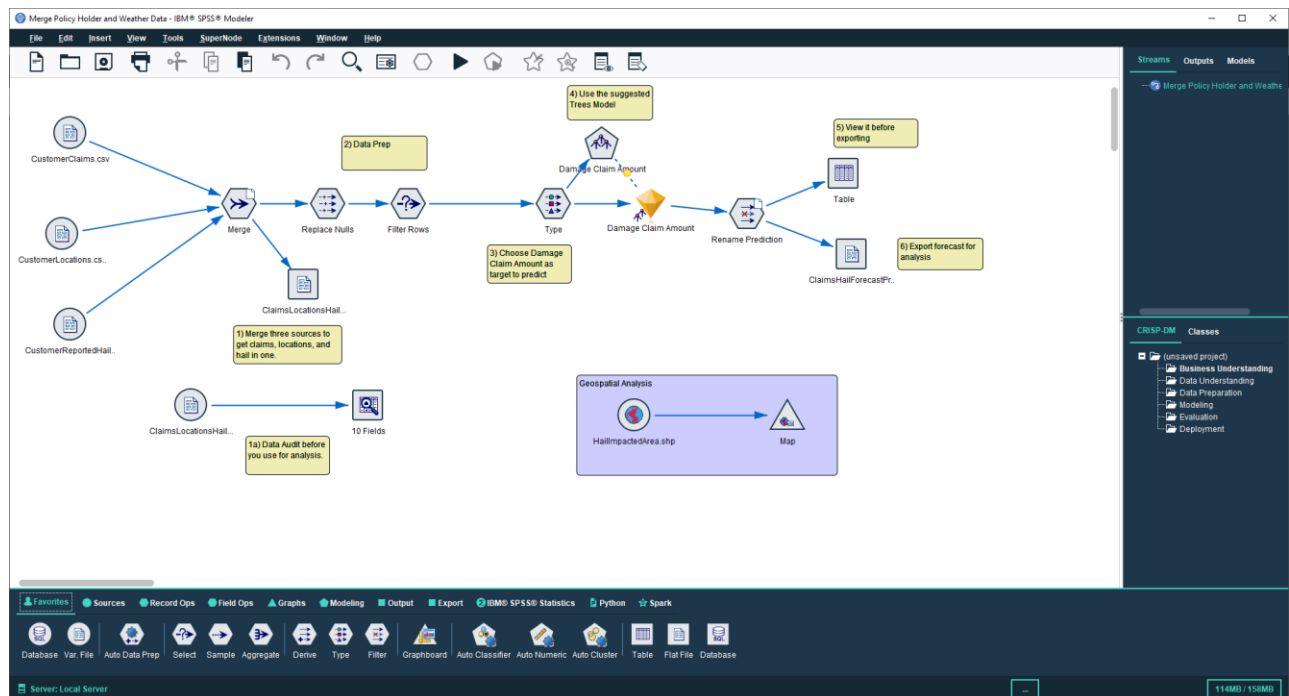


Figure 10 SPSS Modeler Desktop to merge three data sources, prepare the data, build the Extra Trees Regressor Model, prepare it for deployment and deploy to downstream persistent storage for analysis to help others understand the leading predictors and the actual damage amount predicted due to an upcoming severe weather event. Also looking at geospatial analysis within Modeler too.

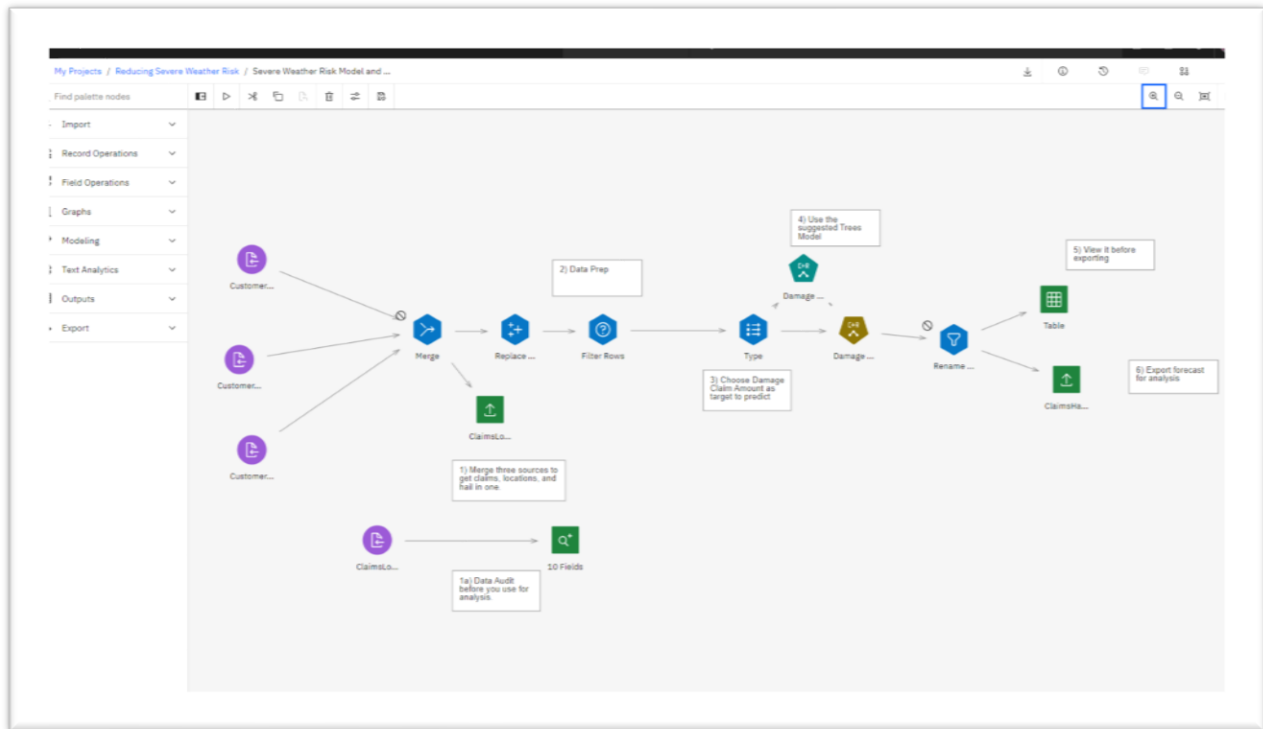


Figure 11 Either import your SPSS Modeler Stream or build from scratch using the web equivalent also part of Cloud Pak For Data.

Once models are built, we promote them to IBM **Watson Machine Learning** model repository to make them available as a live online or batch scoring service.

> Notebooks

> Models

Watson Machine Learning models

Name	Type	Software specification
Hail Risk Modeler Flow Model	spss-modeler_18.1	spss-modeler_18.1
Insurance Hail Risk Auto AI - P1 RandomForestRegressorEstimator	wml-hybrid_0.1	hybrid_0.1

Functions

Name	Type	Software specification	Last modified
You don't have any Functions yet			

> Streams flows

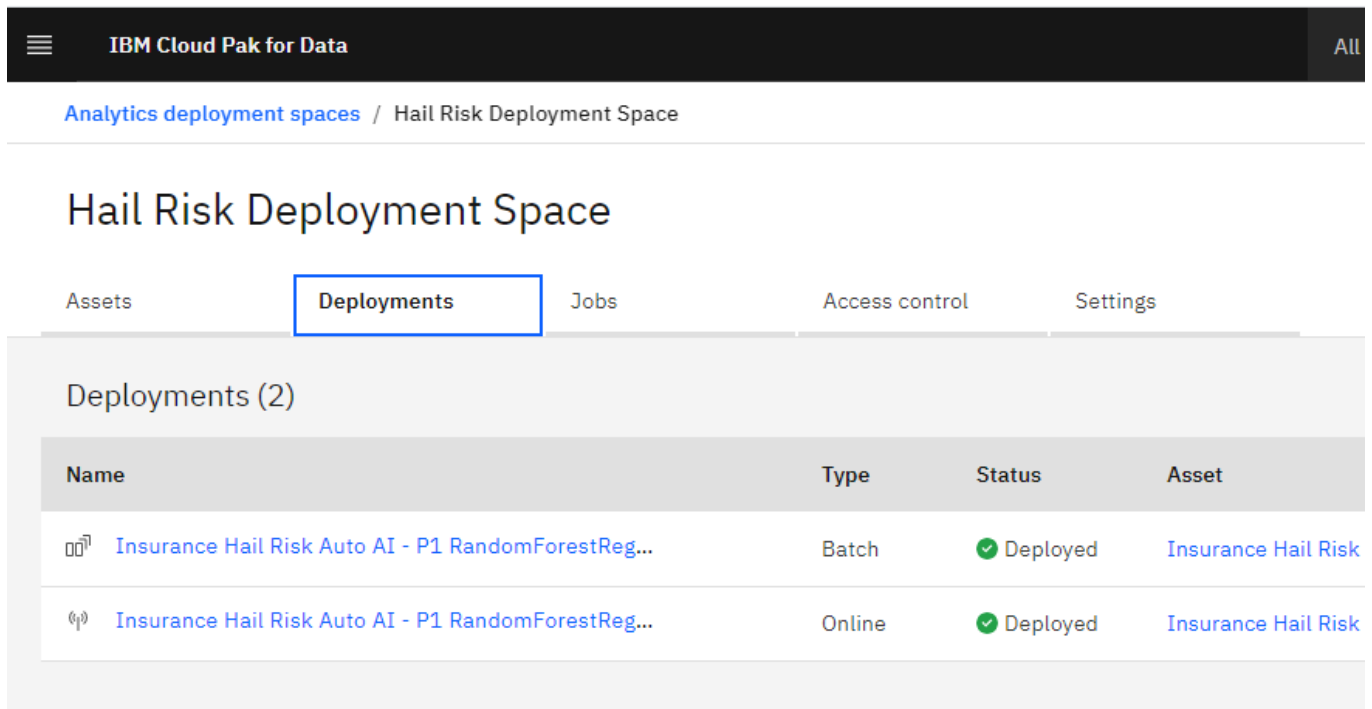


Figure 12 Examples of model deployed. This repository can help manage models and make them available for others to consume and use.

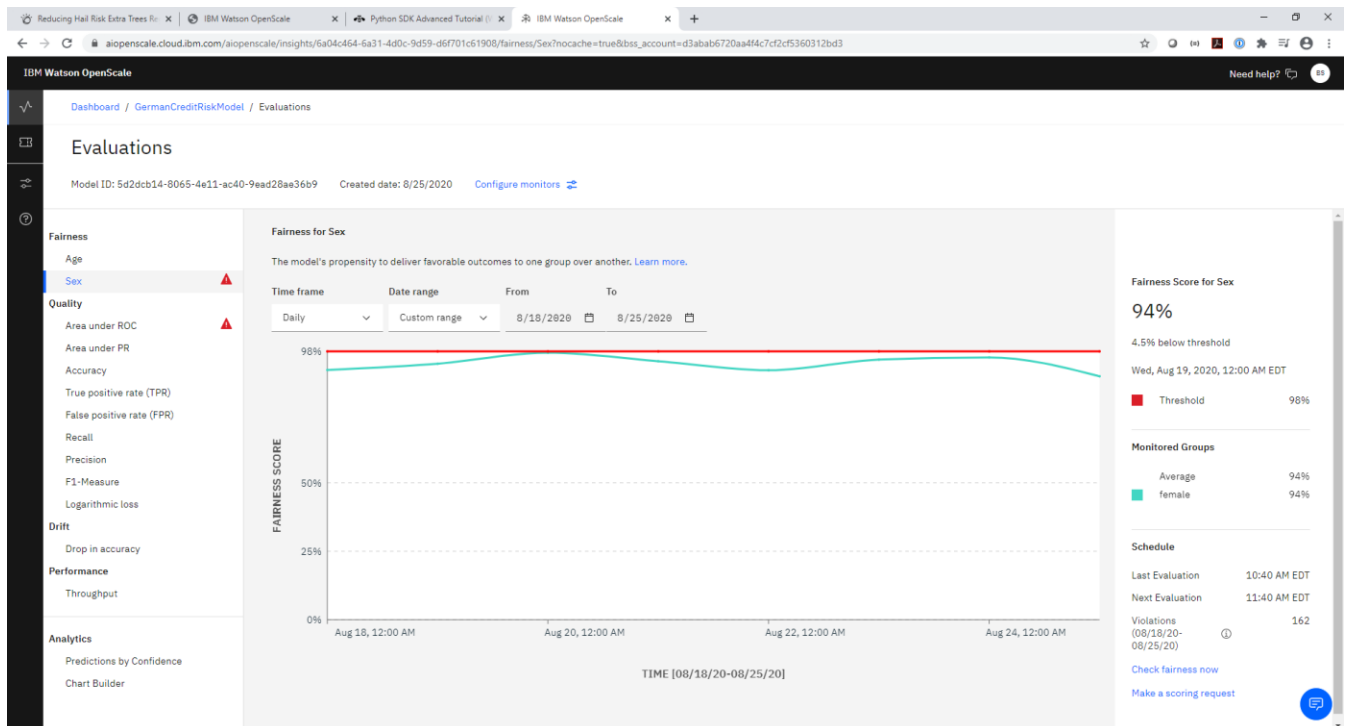


Figure 13 Analyze Model Accuracy, Quality, and Fairness Using IBM Watson OpenScale

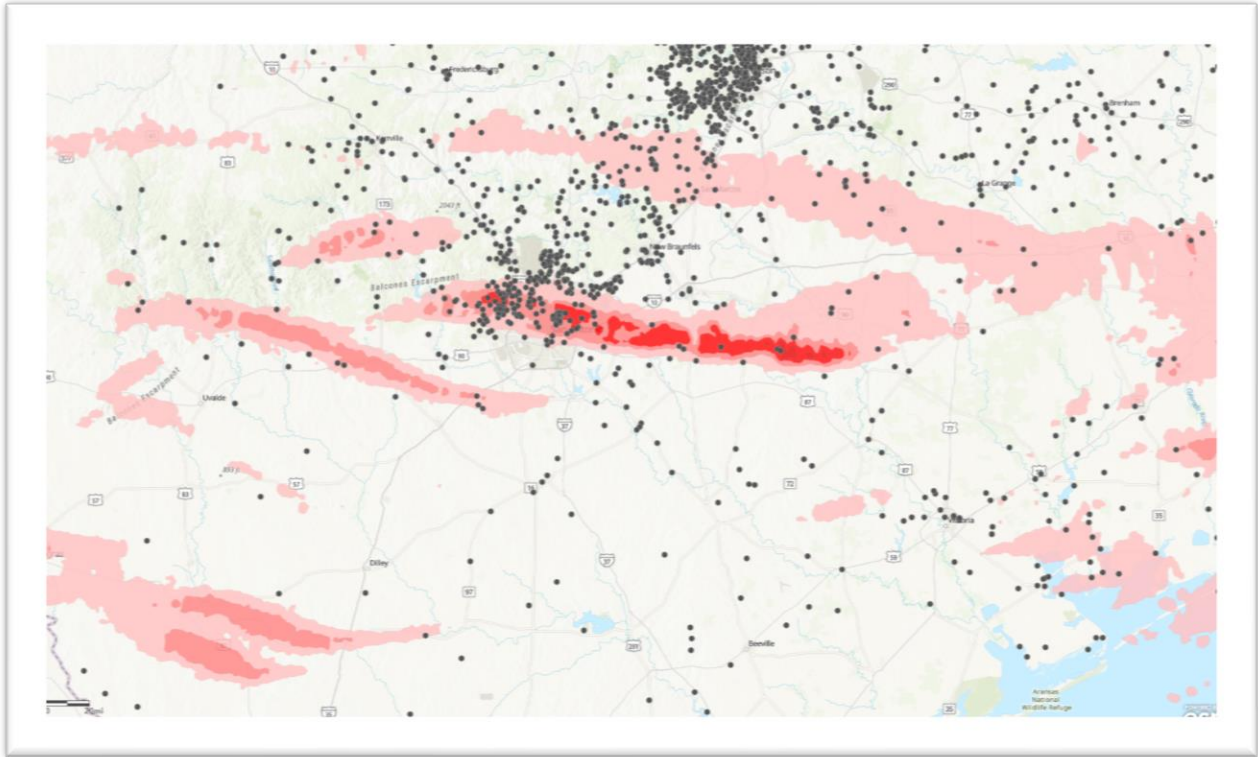


Figure 11 - The Weather Company Accumulated Hail and Damage Claim Locations on ESRI map. A Cloud Pak For Data Python Notebook was used to call the IBM Weather Company Data Severe Weather Package and using the ESRI API for Python push its results to the map on a regular updated schedule. Click on the image for a live example. For more information on using weather data @ IBM go here: <https://www.ibm.com/weather>

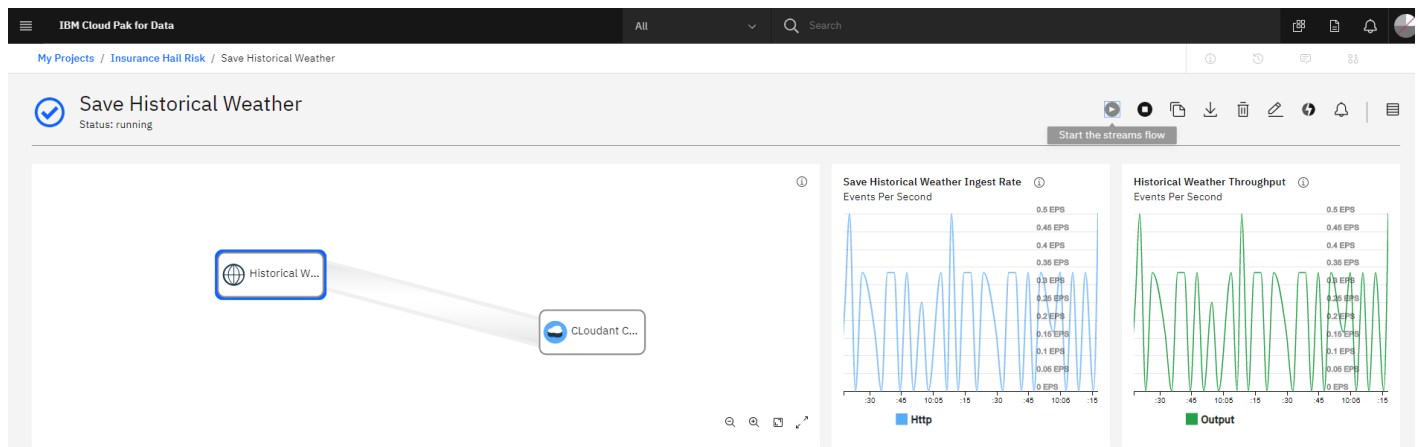


Figure 12 Some insurers prefer to collect severe weather for future analysis. This is an example of saving severe weather to Cloudant on a regular schedule also available using Cloud Pak For Data.

Data Assets Included With The Industry Accelerator

- CSV files
 - **CustomerClaims.csv** – The actual claim amount by customer.
 - **CustomerLocations.csv** – The location of the claim.
 - **CustomerReportedHail.csv** – The hail reported by customer.
 - **ClaimsLocationHail.csv** – Merged file containing claims, location, and hail.
 - **SPSSTreesModelPredicitons.csv** – Output from SPSS Modeler to be analyzed in Analytic Dashboard.
- 1 Analytics Dashboard
 - **Reducing Severe Weather Risk Dashboard.json** – Create a new dashboard using this file.
- 1 SPSS Modeler Stream 18.2.1
 - **Insurance Hail Risk SPSS Desktop Model.str**
 - Use with IBM SPSS Modeler Desktop.
 - Download and try [here](#) or contact bsnyder@us.ibm.com.
- 1 Shapefile
 - **HailImpactedArea.shp**
- 9 Python Notebooks
 - **Reducing Hail Risk Extra Trees Regressor** - Machine Learning Python model using Extra Trees Regressor to predict damage claim amounts with example on promoting to Watson Machine Learning
 - **probabilistic_forecast** - Learn how to call The Weather Company Probabilistic Forecast, used by insurers, and analyze its results in a Python Notebook on CP4D.
 - **Update Hail - Accumulation Maps** - Learn how to call the The Weather Company Severe Weather API and push its results to an ESRI map using the ArcGIS API in a scheduled notebook to update maps used in dashboard.
 - **Update Fire MODIS Thermal Anomalies Maps** - Similar to Update Hail - Accumulation Maps except for a different ESRI map.
 - **Update Gust Wind Speed Maps** - Similar to Update Hail - Accumulation Maps except for a different ESRI map.
 - **Update Sustained Wind Speed Maps** - Similar to Update Hail - Accumulation Maps except for a different ESRI map.
 - **Hail-NA Historical Accumulation - Saved** - Similar to Update Hail - Accumulation Maps except for a different ESRI map.
 - **Update 15 Min Sustain Wind Speed Forecast Maps** - Similar to Update Hail - Accumulation Maps except for a different ESRI map.
 - **Severe Weather Risk Damage Prediction - P1 noteboo** - Example of Python model created using AutoAI to predict damage claim amount.

How To Import The Industry Accelerator?

- 1) Download the .tar.gz file from the [Industry Accelerator Page \(Reducing Severe Weather Risk in Insurance\)](#).
Or go here: <https://ibm.box.com/v/cp4d-insurance-hail-IA>
 - Import **Import-This-Zip-File-Into-CP4D-Reducing-Severe-Weather-Risk.zip** in folder named **1) Import This Project (See #3)**
- 2) Watch [this recording](#).
- 3) Unzip the .tar file and a folder containing what you need will appear.
- 4) Create a new CP4D project using file: **Import-This-Zip-File-Into-CP4D-Reducing-Severe-Weather-Risk.zip**
- 5) Create a new dashboard from file: **Reducing Severe Weather Risk Dashboard.json**
- 6) Create Auto AI using Damage Claim Amount prediction using **SPSSTreesModelPredicitons.csv**
- 7) Examine the SPSS Modeler Stream (optional) using **Insurance Hail Risk SPSS Desktop Model.str**
- 8) Examine CP4D Modeler Flow using file **Insurance Hail Risk SPSS Desktop Model.str**

Note: If you have any questions completing this step, please reach out to bsnyder@us.ibm.com

Prerequisites

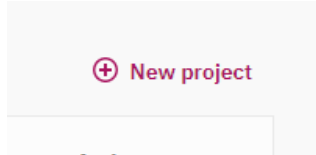
Required services: To use the industry accelerators, you must have the following services available on IBM® Cloud Pak for Data

Service/Software	Required Info
Watson Studio	See Installing Watson Studio . Make sure the Analytic Dashboards, Modeler Flows, Python Notebooks, Auto AI, Watson Machine Learning, Watson Open Scale and Watson Knowledge Catalog are enabled. The default install contains all the services except for Modeler Flows and Cloudant which can easily be turned on.
(Optional) SPSS Modeler Desktop 18.2.x	This is optional if you want to investigate the SPSS Modeler Desktop model. Go here to download a trial or contact bsnyder@us.ibm.com
(Optional) Modeler Streams	Go here to learn more about Modeler Streams. Go here to learn how to enable the Modeler Stream Service. Once enabled, you can import .str files from Modeler Desktop and export .str files to desktop too.

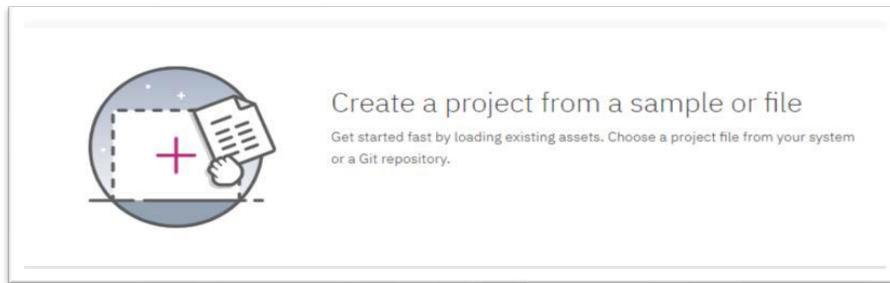
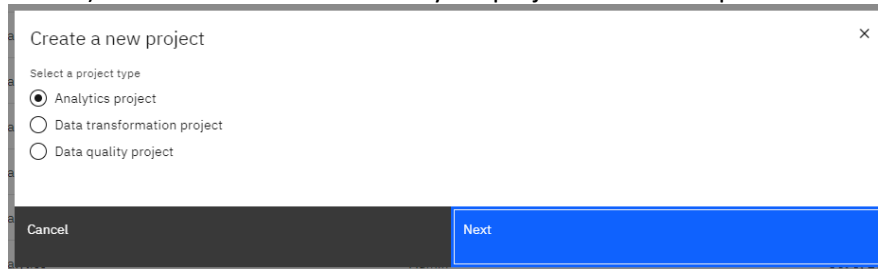
How To Import The Watson Studio Project, Dashboard and Modeler Flow

Import The Project

- 1) In CP4D, Create a New Project



- 2) Choose Create a new Analytics project from a sample or file



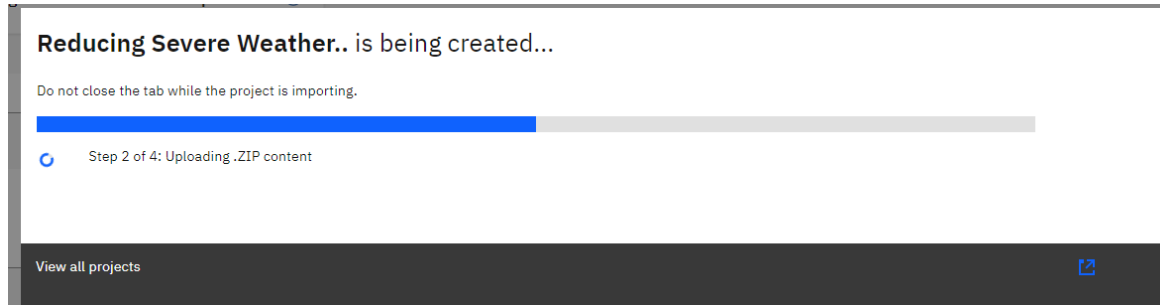
- 3) Select the .zip file, **Import-This-Zip-File-Into-CP4D-Reducing-Severe-Weather-Risk.zip** and populate name and description with **Reducing Severe Weather Risk In Insurance**

A screenshot of the 'Create a project' form. At the top is a header bar with the IBM Cloud Pak for Data logo. Below it, the title 'Create a project' is followed by two tabs: 'From a file' (active) and 'From a Git Repository...'. The main content area has a heading 'Choose a .ZIP file that contains an exported analytics project.' followed by a 'Select file' label with a green checkmark. A file selection box shows 'Import-This-Zip-File-Into-CP4D-Reducing-Severe-Weather-Risk.zip' with a close button. Below this are input fields for 'Name' (containing 'Reducing Severe Weather Risk In Insurance') and 'Description' (containing 'Reducing Severe Weather Risk In Insurance').

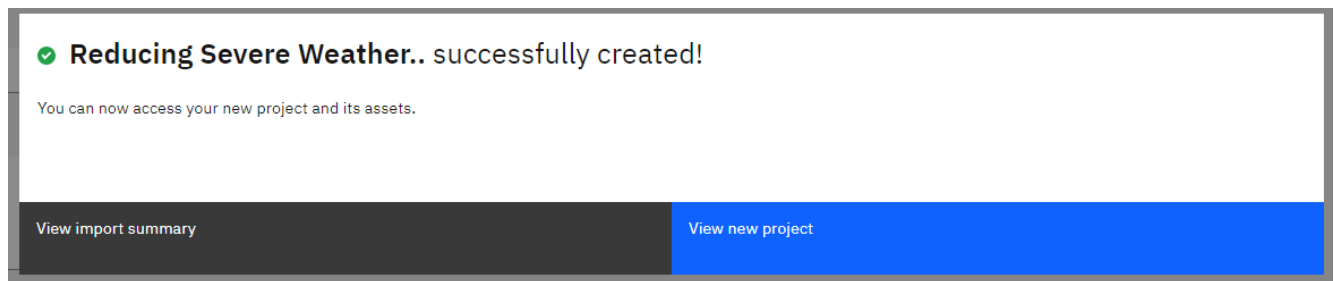
- 4) In the lower right portion of the screen, Click **Create**



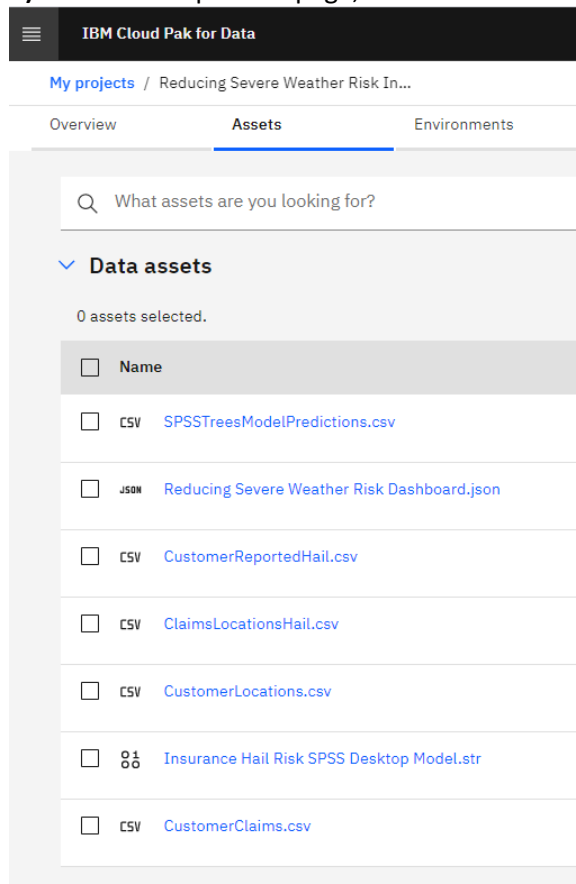
5) You'll see the files import ...



6) After you see the message, **successfully created**, Click **View New Project**



7) At the top of the page, Click on **Assets** to view the files imported successfully.



Import the Analytics Dashboard

- 1) Download **Reducing Severe Weather Risk Dashboard.json** from your Data Assets

▼ Data assets

0 assets selected.

<input type="checkbox"/>	Name	Type	Created by	Last modified	
<input type="checkbox"/>	CSV SPSSTreesModelPredictions.csv	Data Asset	Brian Snyder	Oct 05, 2020, 02:13 PM	
<input type="checkbox"/>	JSON Reducing Severe Weather Risk Dashboard.json	Data Asset	Brian Snyder	Oct 05, 2020, 02:13 PM	⋮
<input type="checkbox"/>	CSV CustomerReportedHail.csv	Data Asset	Brian Snyder	Oct 05, 2020, 02:13 PM	
<input type="checkbox"/>	CSV ClaimsLocationsHail.csv	Data Asset	Brian Snyder	Oct 05, 2020, 02:13 PM	
<input type="checkbox"/>	CSV CustomerLocations.csv	Data Asset	Brian Snyder	Oct 05, 2020, 02:13 PM	

Publish to Catalog

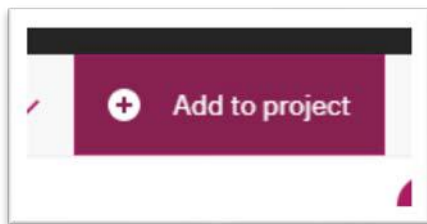
Refine

Download

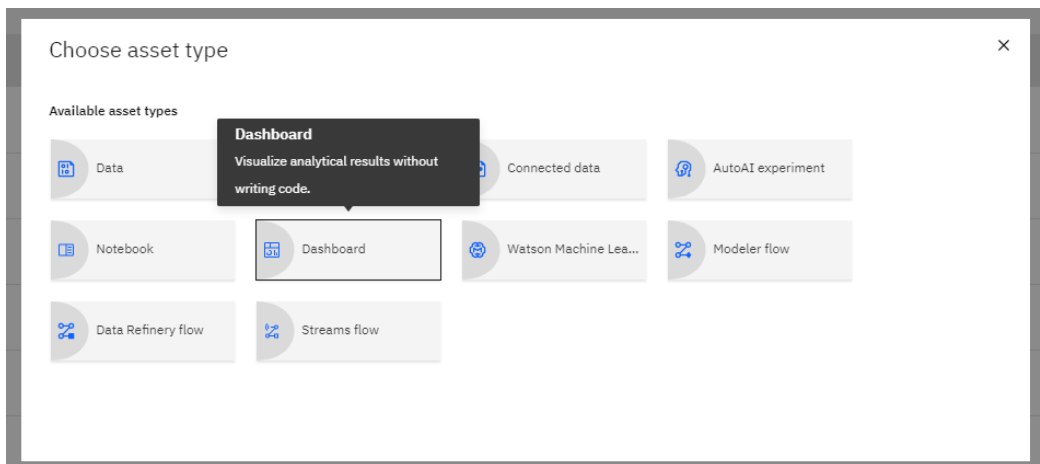
Promote

Remove

- 2) Click **Add to project**



- 3) Select **Dashboard**.



- 4) Choose **From File** using **Reducing Severe Weather Risk Dashboard.json**, use **Reducing Severe Weather Risk Dashboard** for name and description. Click **Create**
- 5) You'll be prompted to re-link the first data asset for **SPSSTreesModelPredictions.csv**. Click **Re-link**.

Missing data asset (1/2)

An associated data set (**SPSSTreesModelPredictions.csv**) couldn't be found. Would you like to re-link the asset now?

Skip

Re-link

- 6) Choose Data Assets > **SPSSTreesModelPredictions.csv** and **Select**

IBM Cloud Pak for Data

Select connection source

Reducing Severe Weather Risk...	Data assets
Assets (2)	Data assets (7)
Connections >	ClaimsLocationsHail.csv
Data assets >	CustomerClaims.csv
	CustomerLocations.csv
	CustomerReportedHail.csv
	Insurance Hail Risk SPSS Desktop...
	Reducing Severe Weather Risk Da...
	SPSSTreesModelPredictions.csv

- 7) Next, you'll be promoted to re-link to **ClaimsLocationsHail.csv** Click **Re-link**

Missing data asset (2/2)

An associated data set (**ClaimsLocationsHail.csv**) couldn't be found. Would you like to re-link the asset now?

Skip

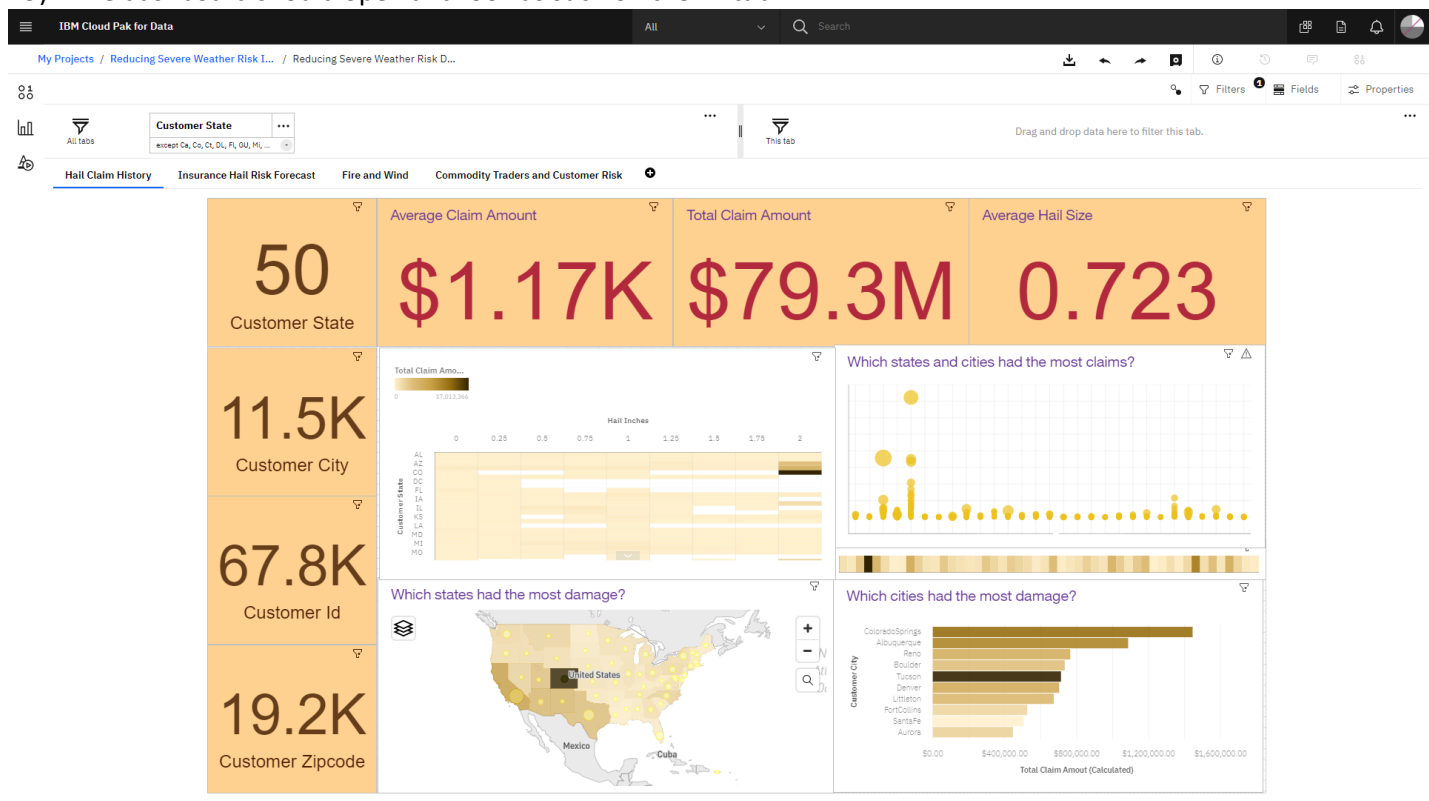
Re-link

- 8) Choose Data Assets > **ClaimsLocationsHail.csv** and click **Select**

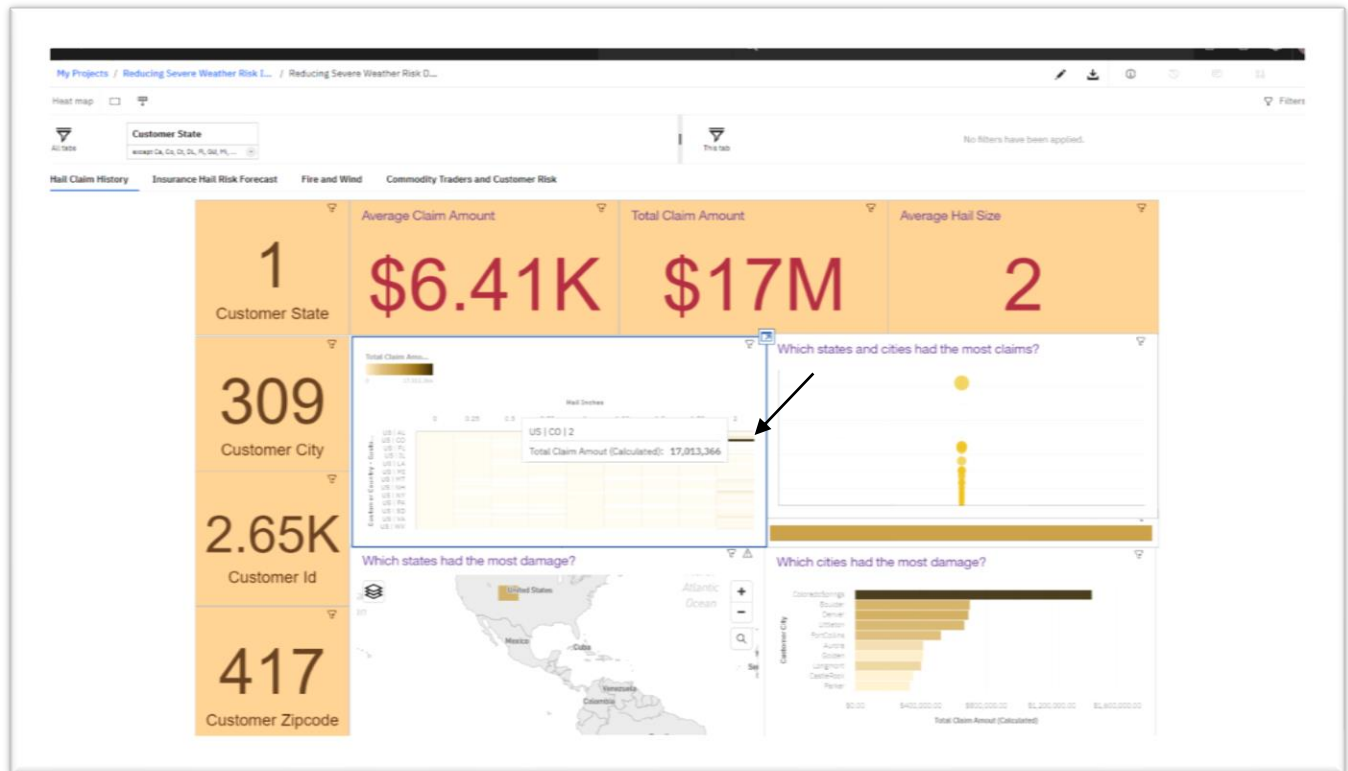
Select connection source

Reducing Severe Weather Risk...	Data assets
Assets (2)	Data assets (7)
Connections	ClaimsLocationsHail.csv
Data assets	CustomerClaims.csv
	CustomerLocations.csv
	CustomerReportedHail.csv
	Insurance Hail Risk SPSS Desktop...
	Reducing Severe Weather Risk Da...
	SPSSTreesModelPredictions.csv

9) The dashboard should open and look as such on the 1st tab:

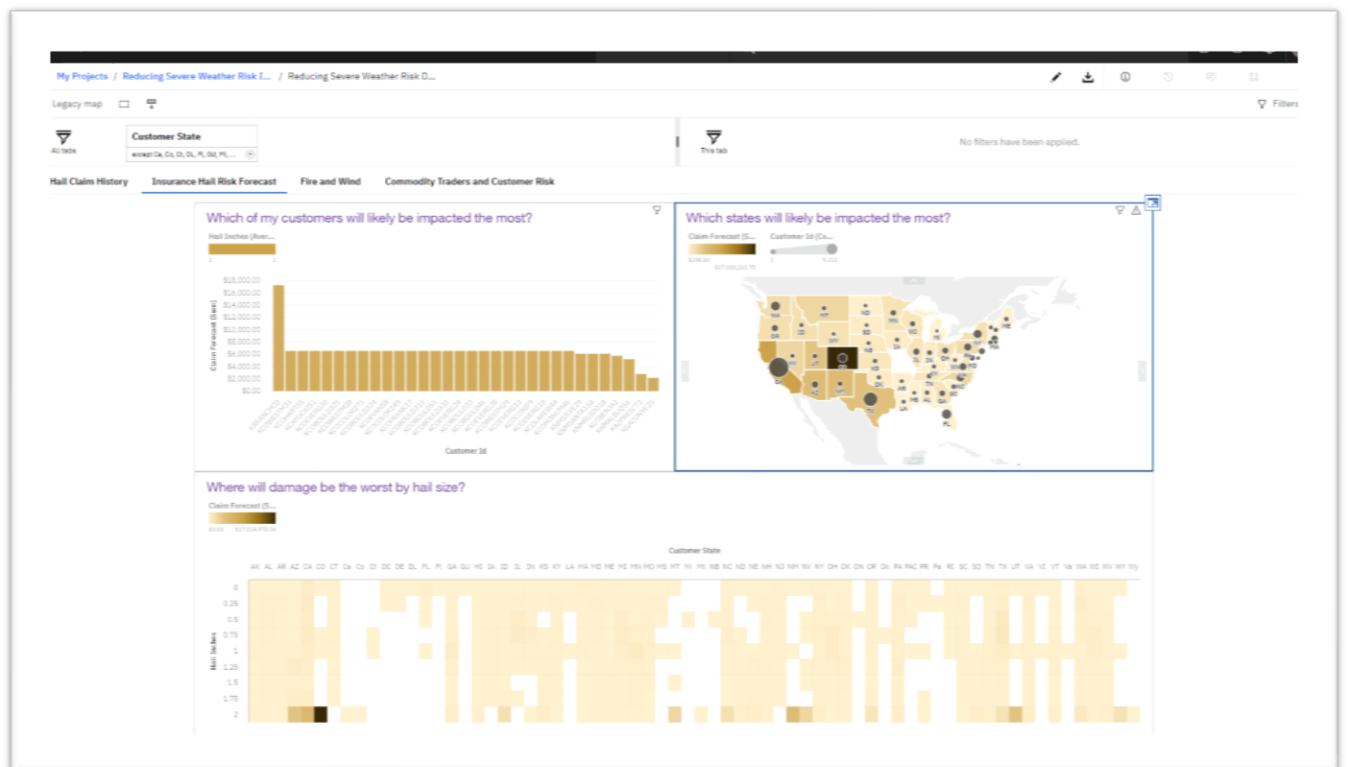


10) The dashboard is interactive. Click on the dark area representing 2" of hail in Colorado. Notice how all the other widgets are focused and filtered on CO.



Note: To remove the focal point, click in an unselected area within the chart.

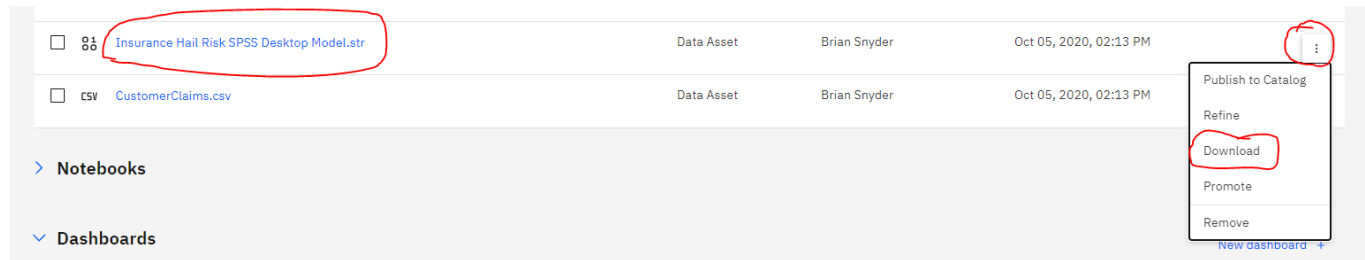
11) Select the 2nd tab, **Insurance Hail Risk Forecast**, highlights where predicted damage will be the worst.



12) The 3rd and 4th Tabs, **Fire and Wind** are **Commodity Traders and Customer Risk** represent ESRI Story Maps that use Cloud Pak For Data to call The IBM Weather Company Weather Data Service and push the geospatial data to an ESRI map on a schedule for near real time updates. Scroll down to see the upcoming severe wind, fire and long term seasonal forecast on each tab. View the imported Watson Studio notebooks to view the to pull and push code used.

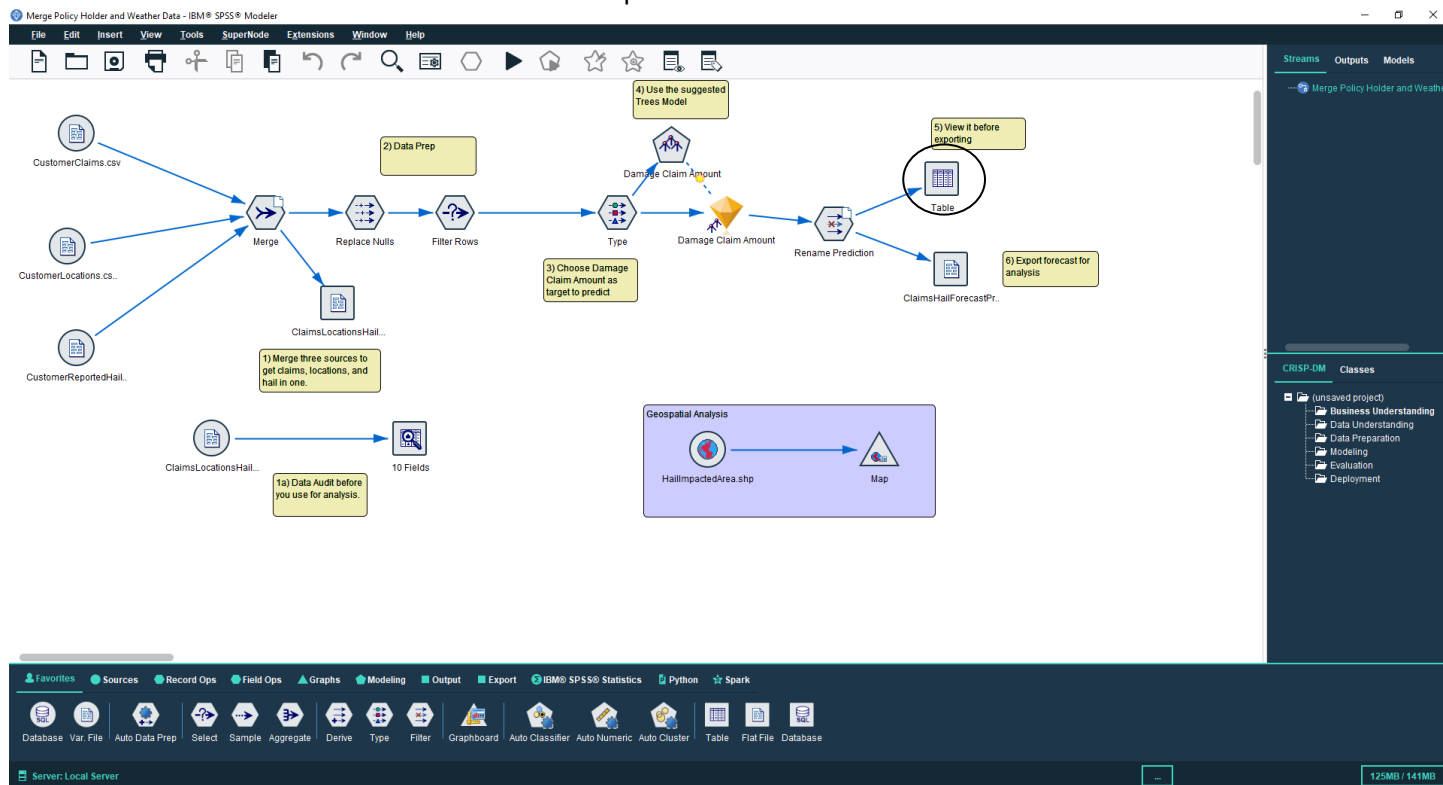
Importing the SPSS Modeler Stream and CP4D Modeler Flow

- 13) To download the SPSS Modeler File, *Insurance Hail Risk SPSS Desktop Model.str*, go to your Data Assets and choose to download it.



- 14) If you have SPSS Modeler Desktop 18.2.x or greater, you can explore the model (*Insurance Hail Risk SPSS Desktop Model.str*) on your own.

Here's the stream in SPSS Modeler Desktop.



Step 1 is the first tab of the dashboard showing historical hail damage and claim amounts.. Step 6 is the 2nd tab of the dashboard showing predicted claim damage.

- 15) If you've opened in SPSS Modeler, then right click and Run the Table Node to view the results used in the 2nd tab of the dashboard.

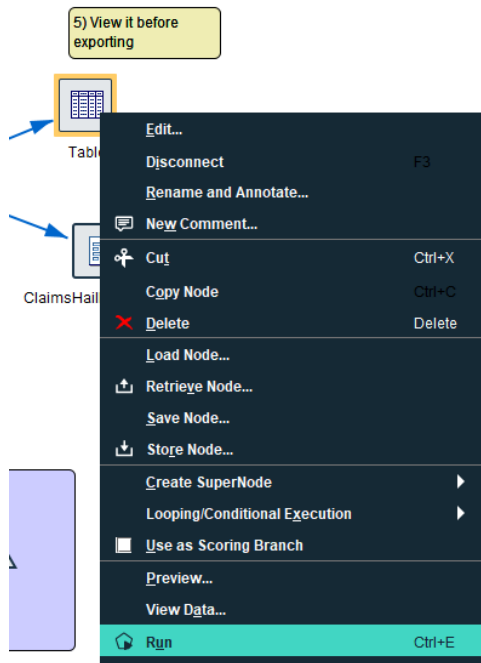


Table (10 fields, 68,584 records)

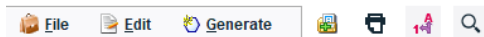
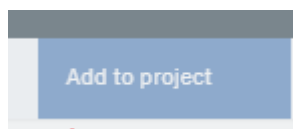


Table Annotations

	Customer Id	Hail Inches	Customer Latitude	Customer Longitude	Customer Zipcode	Customer City	Customer State	Customer Country	Customer Since	Predicted Claim Amount
1	1	0.500	61.660	-149.330	99654	Wasilla	AK	US	2009	488.030
2	2	1.500	68.840	-148.830	999	Alaska	AK	US	2013	999.570
3	3	2.000	68.630	-149.600	999	Alaska	AK	US	2014	2153.515
4	4	0.000	70.430	-150.410	999	Alaska	AK	US	2016	115.239
5	5	1.000	61.180	-149.640	999	Alaska	AK	US	2016	1021.811
6	6	0.000	61.230	-149.890	99501	Anchorage	AK	US	2012	115.239
7	7	0.000	61.220	-149.840	99501	Anchorage	AK	US	2012	115.239
8	8	0.000	61.100	-149.830	99516	Anchorage	AK	US	2012	115.239
9	9	0.000	61.110	-149.890	99515	Anchorage	AK	US	2012	115.239
10	10	1.750	61.070	-149.740	99516	Anchorage	AK	US	2012	2153.515
11	11	0.000	61.210	-149.920	99501	Anchorage	AK	US	2013	115.239
12	12	0.000	61.220	-149.760	99504	Anchorage	AK	US	2013	115.239
13	13	2.000	61.270	-149.480	99577	Anchorage	AK	US	2013	2153.515
14	14	1.500	61.120	-149.720	99507	Anchorage	AK	US	2004	999.570
15	15	0.500	61.400	-149.430	99567	Anchorage	AK	US	2013	488.030
16	16	0.000	61.390	-149.510	99567	Anchorage	AK	US	2013	115.239
17	17	0.000	61.130	-149.840	99507	Anchorage	AK	US	2014	115.239
18	18	0.000	61.180	-149.770	99504	Anchorage	AK	US	2014	115.239
19	19	1.000	61.130	-149.720	99507	Anchorage	AK	US	2014	999.570
20	20	1.000	61.210	-149.820	99508	Anchorage	AK	US	2014	999.570
21	21	1.000	61.330	-149.570	99577	Anchorage	AK	US	2014	999.570

Import as a Modeler Stream in CP4D.


16) Choose **Add to project, Modeler Flow, From File**. Use *Insurance Hail Risk SPSS Desktop Model.str*. Name it *Hail Risk Modeler Flow* and click **Create**.





Choose asset type





Available asset types


 Data


 Connection


 Connected data


 Notebook

 Dashboard

 Watson Machine Lea...

 Modeler flow

 Data Refinery flow

 Streams flow

Modeler flow
Create a graphical representation of a data model or a neural network design.

IBM Cloud Pak for Data

New modeler flow

New

From File

From Example

Name

Hail Risk Modeler Flow

Description (optional)

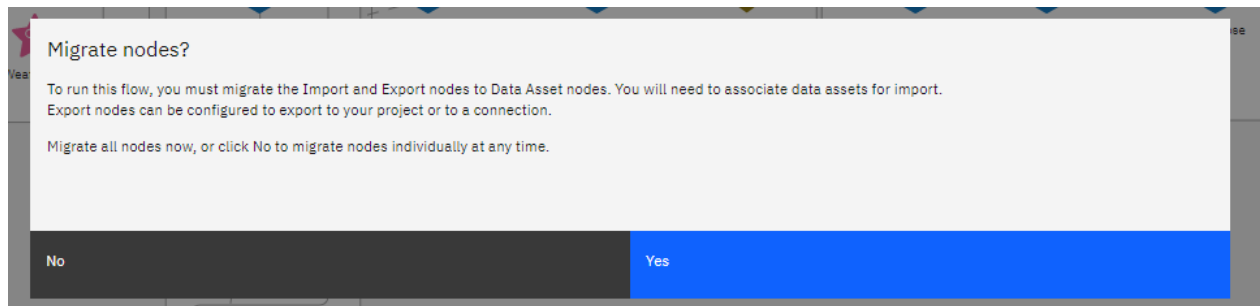
Hail Risk Modeler Flow

Upload flow file

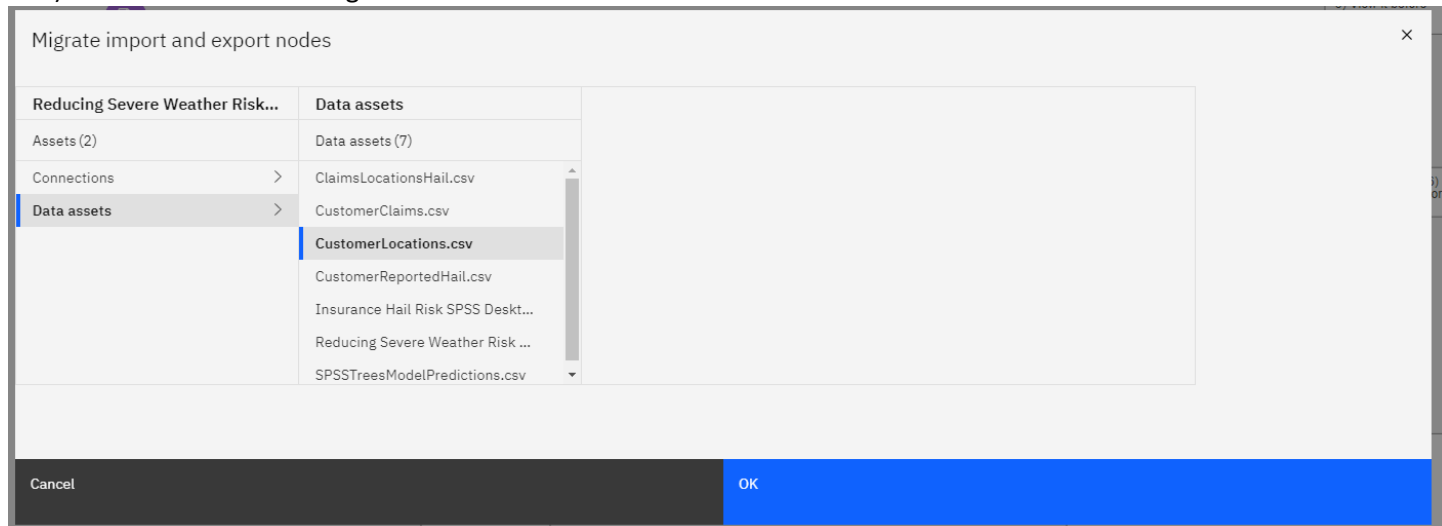
Drag and drop an SPSS Modeler flow file here or [browse](#) your local device to select a file.

Insurance Hail Risk SPSS Desktop Model.str

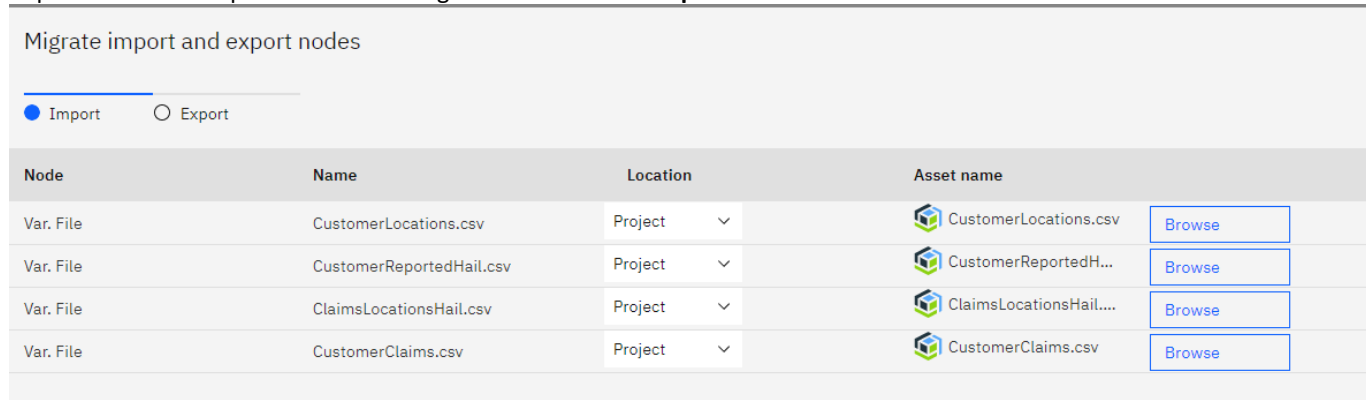
17) You'll be prompted to Migrate nodes? Click **Yes**



18) Click **Browse** for to assign Customer Locations to Data Assets > CustomerLocations.csv. Click **OK**

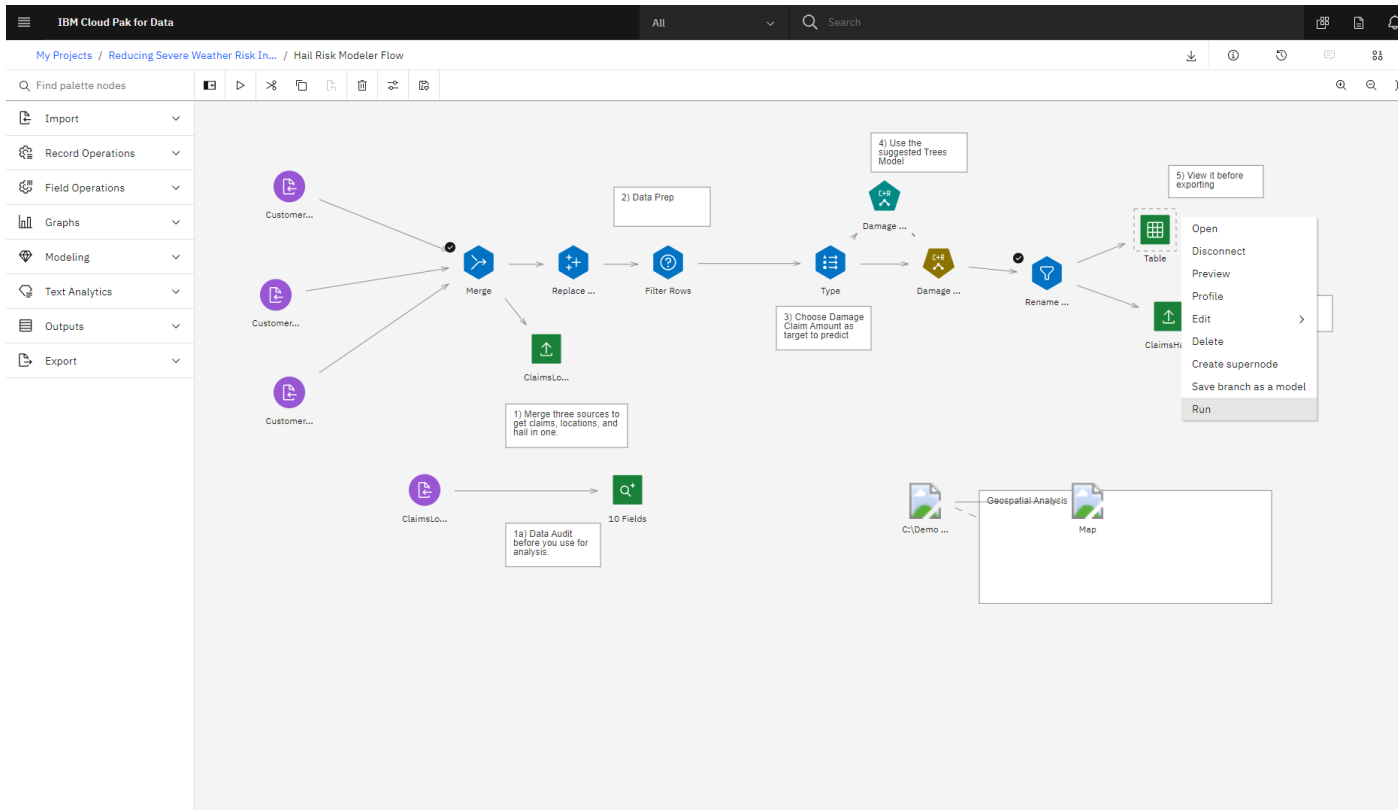


19) Repeat the same steps for the remaining 3 files. Click **Go to Export**



20) Map the 2 files to the appropriate Exports and click **Migrate**

21) Click and run #5 to verify the execution. Open the Table to view its results.



IBM Cloud Pak for Data

All

Search

My Projects / Reducing Severe Weather Risk In... / Hail Risk Modeler Flow / Table (10 fields, 68,584 records)

← Return to flow

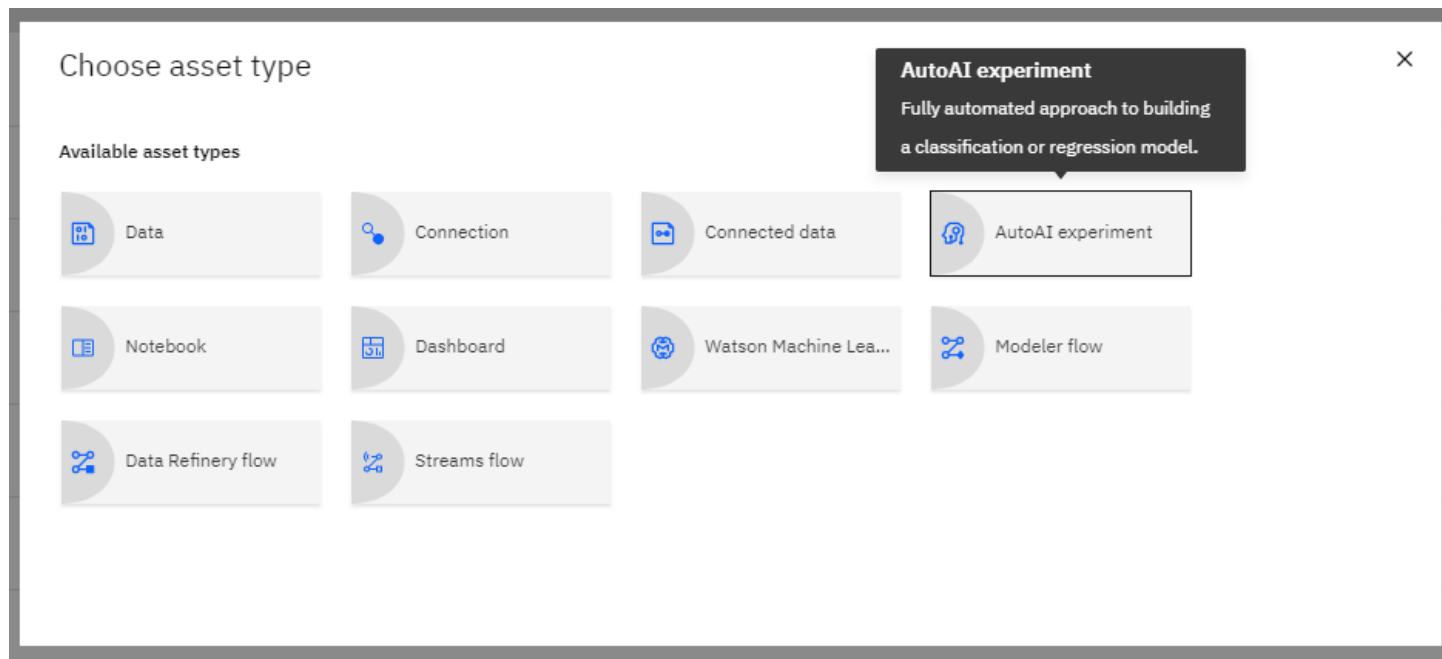
View Output: Table (10 fields, 68,584 records)

Customer Id	Hail Inches	Customer Latitude	Customer Longitude	Customer Zipcode	Customer City	Customer State	Customer Country	Customer Since	Predicted Claim Amount
1	0.500	61.660	-149.330	99654	Wasilla	AK	US	2009	488.030
2	1.500	68.840	-148.830	999	Alaska	AK	US	2013	999.570
3	2.000	68.630	-149.600	999	Alaska	AK	US	2014	2153.515
4	0.000	70.430	-150.410	999	Alaska	AK	US	2016	115.239
5	1.000	61.180	-149.640	999	Alaska	AK	US	2016	1021.811
6	0.000	61.230	-149.890	99501	Anchorage	AK	US	2012	115.239
7	0.000	61.220	-149.840	99501	Anchorage	AK	US	2012	115.239
8	0.000	61.100	-149.830	99516	Anchorage	AK	US	2012	115.239
9	0.000	61.110	-149.890	99515	Anchorage	AK	US	2012	115.239
10	1.750	61.070	-149.740	99516	Anchorage	AK	US	2012	2153.515
11	0.000	61.210	-149.920	99501	Anchorage	AK	US	2013	115.239
12	0.000	61.220	-149.760	99504	Anchorage	AK	US	2013	115.239
13	2.000	61.270	-149.480	99577	Anchorage	AK	US	2013	2153.515
14	1.500	61.120	-149.720	99507	Anchorage	AK	US	2004	999.570
15	0.500	61.400	-149.430	99567	Anchorage	AK	US	2013	488.030
16	0.000	61.390	-149.510	99567	Anchorage	AK	US	2013	115.239

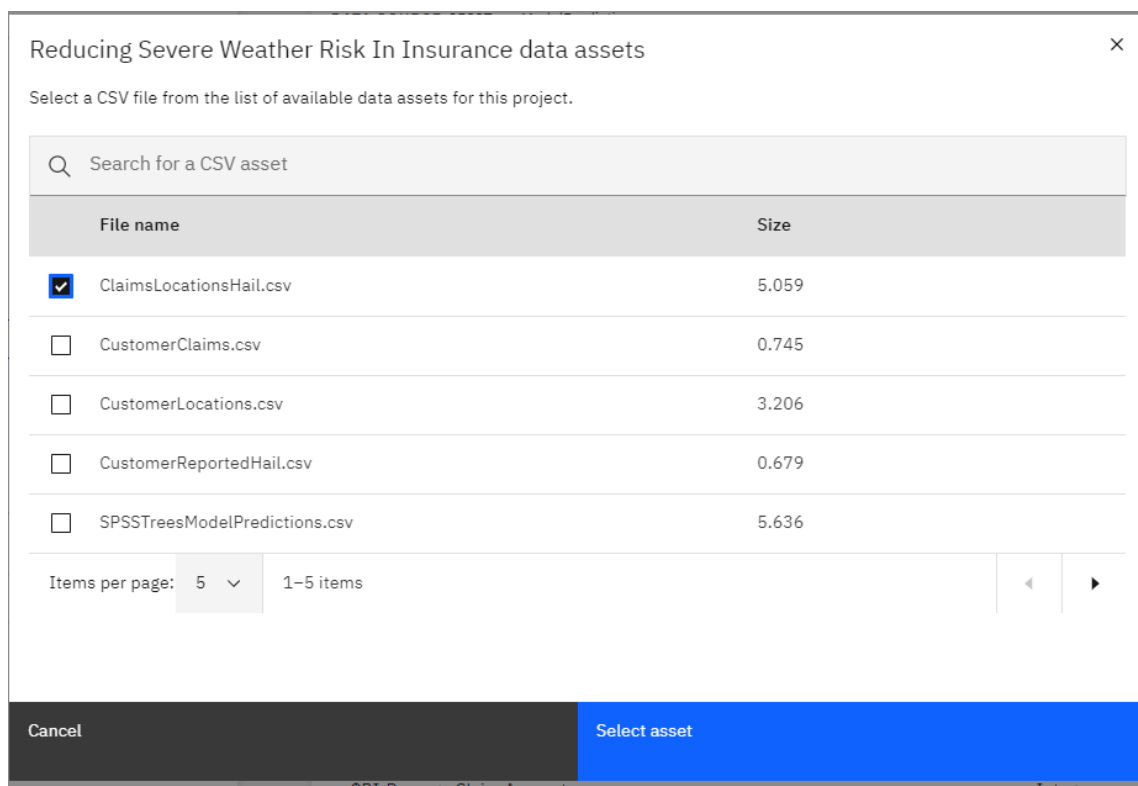
Items per page: 200 1–200 of 68584 items 1 of 343 pages

Using AutoAI For Model Selection And Analysts That Want Data Science But Aren't Data Scientists

- 1) Choose Add To Project > AutoAI Experiment. Name it *Severe Weather Risk Damage Prediction* and click **Create**



- 2) Choose the file named ***ClaimsLocationsHail.csv*** and choose **Select Asset**



- 3) Choose Damage Claim amount as the field we're trying to predict. Notice how it's smart enough to choose regression models to analyze.

Configure AutoAI experiment
Severe Weather Risk Damage Prediction

Add data source

Add data source
Drop a .csv file here or [browse](#) for a file to upload. Maximum file size is 1 GB.
— OR —
[Select from project](#)

Data source name	Size (MB)	Columns
ClaimsLocationsHail.csv	5.059	10

Select prediction column
DATA SOURCE: ClaimsLocationsHail.csv

Column name	Type
Customer Id	Integer
Hail Inches	Decimal
Customer Latitude	Decimal
Customer Longitude	Decimal
Customer Zipcode	Integer
Customer City	String
Customer State	String
Customer Country	String
Customer Since	Integer
Damage Claim Amount	Integer

Prediction column: Damage Claim Amount

PREDICTION TYPE: Regression
OPTIMIZED METRIC: RMSE

Experiment settings [Run experiment](#)

4) Click **Experiment Settings**. Adjust the slider to 85% testing

Experiment settings

Data source settings

PREDICTION COLUMN: Damage Claim Amount
COLUMN DATA TYPE: Integer
DATA SOURCE: ClaimsLocationsHail.csv

Subsample
For a large data set, use a subset of data to train the experiment. This speeds up results but may affect accuracy.
☐ Subsample rows

Training data split
You can optionally adjust the percentage of your data source to use for creating, optimizing, and validating pipelines. Only recommended for large data sets to avoid decreasing the quality of the pipelines.
85%
Training data split: 85% — 3 folds | Holdout data split: 15%

Select columns to include
Select columns with data that support the prediction column. Included columns: 10 / 10

Column name	Type
<input checked="" type="checkbox"/> Customer Id	Integer
<input checked="" type="checkbox"/> Hail Inches	Decimal
<input checked="" type="checkbox"/> Customer Latitude	Decimal
<input checked="" type="checkbox"/> Customer Longitude	Decimal
<input checked="" type="checkbox"/> Customer Zipcode	Integer

Cancel Save settings

5) Click **Prediction** to see the list of algorithms to test, and which metric it will be optimized as well as the Runtime settings to scale for larger models on demand.

The screenshot displays the IBM Cloud Pak for Data interface, specifically the 'Prediction settings' and 'Runtime settings' tabs. The 'Prediction settings' tab is active, showing the following configuration:

- Prediction Column:** Damage Claim Amount
- Column Data Type:** Integer
- Data Source:** ClaimsLocationsHail.csv
- Prediction type:** Regression (selected)
- Optimized metric:** Root Mean Squared Error (RMSE) (Recommended)
- Algorithms to test:** 8 / 8

The 'Runtime settings' tab is also visible, showing the following configuration:

- Experiment details:** Initial model tuning iterations: 25, Feature engineering iterations: 60, Final model tuning iterations: 50
- Compute configuration:** 4 vCPU and 16 GB RAM (selected)

- 6) Click **Save Settings**
- 7) Click **Run Experiment**. Note: this may take a few moments. You need not wait for it to complete. It's not a problem if you close your browser and come back later.

For questions and inquiries please reach out to Brian Snyder, bsnyder@us.ibm.com