

# Financial Markets Customer Attrition Prediction

---

The Financial Markets Customer Attrition Prediction accelerator provides a set of sample data science assets, a structured glossary of business terms, that help you to predict which customers will leave.

**Tip:** Download the PDF of these instructions from the Data assets section on the **Assets** page so you can keep these instructions open while you work.

- [Instructions](#)
- [Sample data assets](#)
- [Notebooks](#)
- [R Shiny dashboard](#)
- [Sample business glossary](#)

## Instructions

---

Follow these steps to implement the industry accelerator:

1. Create a deployment space, which you'll need when you run the notebooks. From the navigation menu, select **Deployments > View All Spaces > New deployment space** and then complete these steps:
  1. Enter the name **Customer Attrition Space**
  2. Select your Watson Machine Learning service. If you don't have one, from the **Select machine learning service** pull-down menu, select **Create a new machine learning service**, provision one, and then click the **Create** button.
  3. Click **Create** to create the deployment space.
2. Navigate back to the accelerator project, select the **Assets** tab and scroll to the **Notebooks** section.
3. Edit the **1-data-preprocessing** notebook by clicking the edit icon that looks like a tiny pencil next to the notebook name. This notebook loads the data and creates and saves the `attrition_prep.py` script to prepare and clean data for model training. It also analyses correlations in the data set. Follow the instructions in the notebook to configure and step through running it.
4. Edit and run the **2-model-training** notebook. This notebook transforms the data, builds machine learning models, and deploys a model.
5. Edit and run the **3-model-scoring** notebook. This notebook deploys the data assets and the model scoring function.
6. Run the dashboard from RStudio console by completing these steps:
  1. Download the `customer-attrition-prediction-analytics-dashboard.zip` file from the Data assets section of the **Assets** page. If you don't see the file, click **View All** to display the full list of assets.
  2. Click **Launch IDE > RStudio** on the menu bar.
  3. In the **Files** pane, select the **Upload** toolbar button and upload the `customer-attrition-prediction-analytics-dashboard.zip` file into RStudio.
  4. Navigate to `customer-attrition-prediction-analytics-dashboard`, select the `app.R` file, and click the **Run App** toolbar button to launch the dashboard. If you see a warning message that certain packages are not installed, you can ignore it because the packages will be installed first time you run the app.

5. Once the app has launched, you can perform model scoring in real time by entering your API key and selecting your region on the Client View tab.
7. 10Optional. You can import the glossary of business terms into Watson Knowledge Catalog to get started on data governance. With the **Lite plan** only 5 Business terms can be imported. You must have [the permission to create governance artifacts](#).
  1. Download the customer-attribution-prediction-glossary-categories.csv and customer-attribution-prediction-glossary-terms.csv files from the Data assets section of the **Assets** page.
  2. Navigate to **Governance > Categories**.
  3. Click **Add Category > Import From File**.
  4. Import the customer-attribution-prediction-glossary-categories.csv file. Select **Replace all values** as your merge option.
  5. Navigate to **Governance > Business Terms**.
  6. Click **Add Business Term > Import From File**.
  7. Import the customer-attribution-prediction-glossary-terms.csv file. Select **Replace all values** as your merge option.
  8. Once the Import completes successfully, click on **Go to task** and then click **Publish** in the next page.
  9. Navigate to **Governance > Categories > Industry Accelerators** to explore the business terms.

**Note:** Importing the accelerator and running the notebooks and dashboard consumes approximately 1.5 - 2 Watson Studio CUH and 3 - 3.5 Watson Machine Learning CUH.

## Sample data assets

---

These sample data files that act as dimensional and fact tables are included in the project on the **Assets** page:

- customer.csv: Customer Data, Demographic data, Temporal data.
- account.csv: Account type and Account Information Data, Investment Information, Temporal data.
- customer\_summary.csv: Detailed Customer Transaction Data, Business Metrics, Investment and Income Stats.
- customer\_history.csv: A join of the previous three data sets based on customer.customer\_id, customer.effective\_date, customer\_summary.customer\_id, customer\_summary.end\_date, account.primary\_customer\_id, account.open\_date and account.close\_date columns. This data set is used as raw data input for the data preparation in the **1-data-preprocessing** notebook. See the CUSTOMER\_HISTORY\_VIEW.sql file for the SQL query used to merge the tables. The resulting data is filtered, transformed, and aggregated to contain one record per customer, so you can use it for modelling.
- account\_summary.csv: Related to the account.csv data set. This data set also contains account related data and investment data. This information is only used to depict financial statistics of the customers in the analytics dashboard.

## Notebooks

---

Follow the instructions in the first few cells of the notebook to configure the project token and API key. Then run the notebooks step-by-step.

- **1-data-preprocessing:** This notebook performs the following functions:
  - Load data

- Create and save script `attrition_prep.py` to prepare and clean data for model training
- Analyze correlations
- **2-model-training:** This notebook performs the following functions:
  - Build ML models
  - Analyze and visualize the data
  - Select the best performing ML model and save it
  - Store the model in the deployment space and deploy the model
- **3-model-scoring :** This notebook performs the following functions:
  - Get the deployment space and deployments
  - Deploy the data assets
  - Create and deploy a pipeline function for model scoring
  - Predict customer attrition

## R Shiny dashboard

---

The R Shiny dashboard displays product statistics and customer summaries, and scores new data. The dashboard has the following tabs:

- **Dashboard View:** Shows top action clients, monthly customer attrition, and customer attrition risk level.
- **Client View:** Targets individual client information, depicts the top business metrics and account details, provides an option to run the model scoring web service, predicts customer attrition, and visualizes the influential factors and data fields.

## Sample business glossary for use with Watson Knowledge Catalog

---

Optionally, you can import the glossary of business terms into Watson Knowledge Catalog to get started on data governance.

The `customer-attribution-prediction-glossary-categories.csv` file defines the main and sub categories for the business terms.

The `customer-attribution-prediction-glossary-terms.csv` file defines the business terms, category of the business terms and their Related Terms/Part of Terms, if applicable.

**Note:** This optional step requires the Watson Knowledge Catalog service instance.

## Terms and Conditions

---

This project contains Sample Materials, provided under this [license](#).

Licensed Materials - Property of IBM.

© Copyright IBM Corp. 2019, 2021. All Rights Reserved.

US Government Users Restricted Rights - Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.