# PRACTICAL INTRODUCTION TO THE IBM DATA SCIENCE EXPERIENCE PLATFORM

## SECTION 1: Loading data in Watson Analytics

This Section requires having Watson Analytics properly setup

Non-IBM users can sign up for a Free Trial of Watson Analytics at the following link (for the US-English version): https://www.ibm.com/analytics/watson-analytics/us-en/

1. Upon opening Watson Analytics, click the "New data" button.
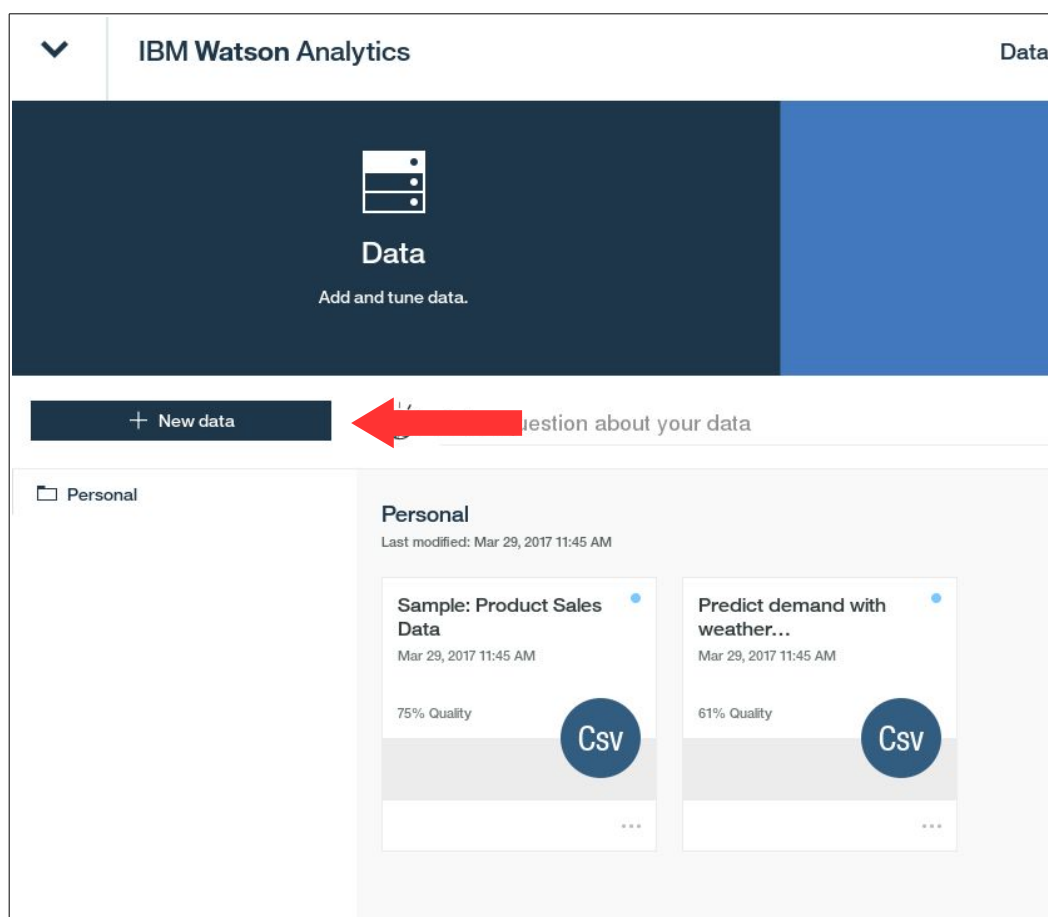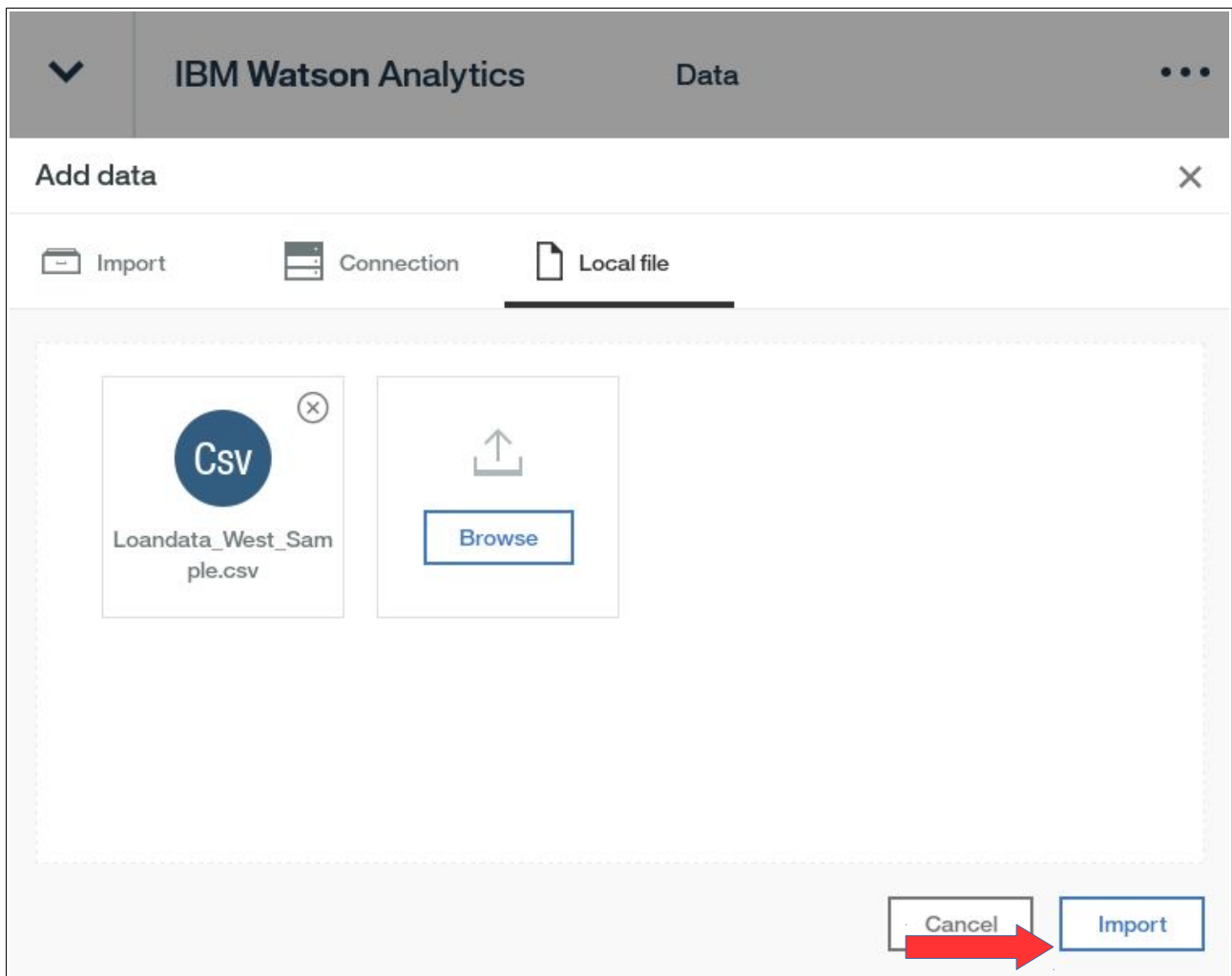

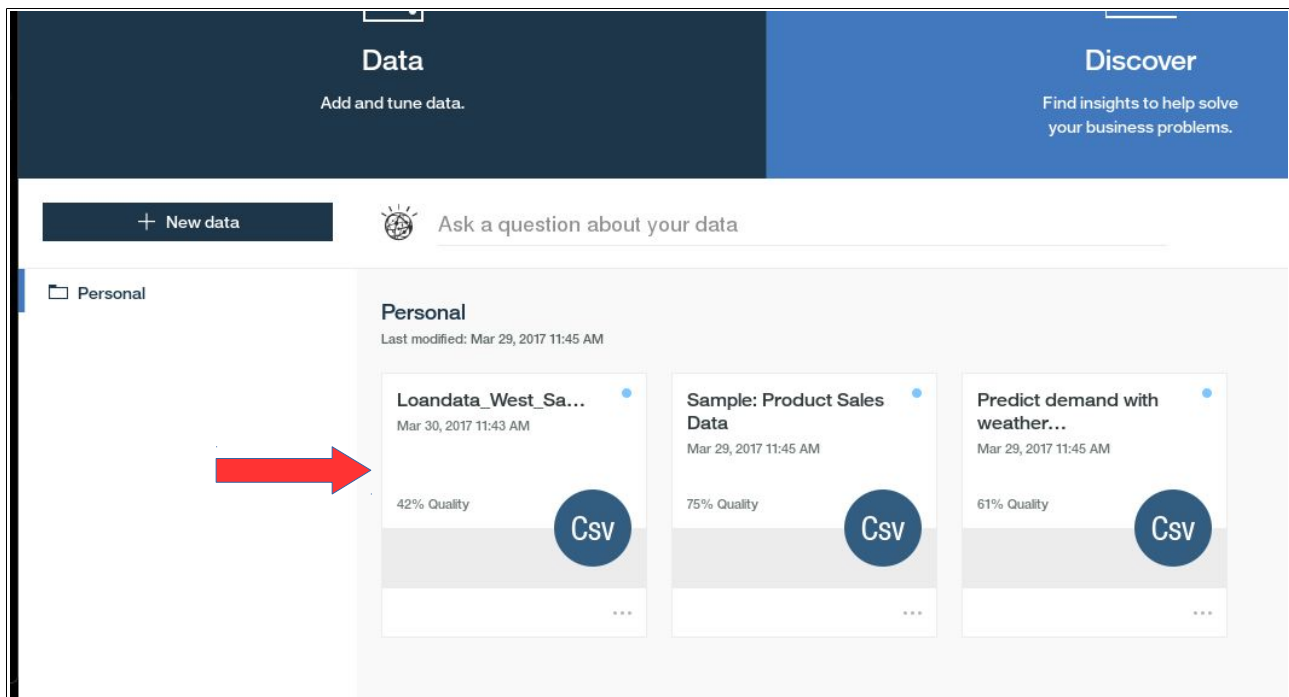2. At the next screen click on the "Local file" button.
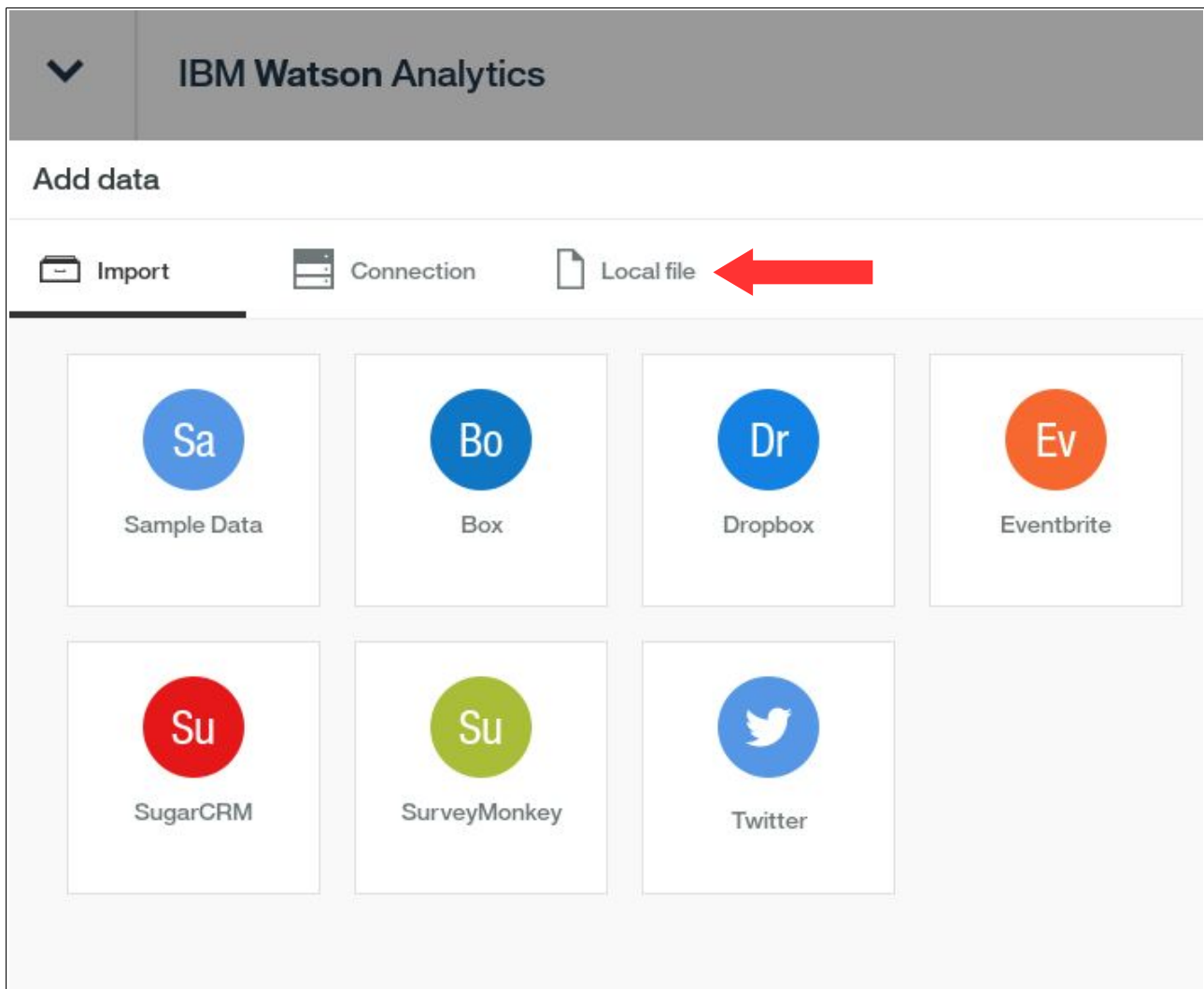
3. Click "Browse", navigate to the location where you stored the Loandata_West_Sample.csv file and select it.

4. Click the "Import" button.

You should now be back at the home screen, where the Loandata_West_Sample file should be listed under your Personal data folder.

Important Note

For demonstration purposes, the data set you are using in this lab is a SMALL subset of an originally much larger data set which was used to generate the documented steps below. Certain results and insights may therefore be different from what is shown in subsequent steps.

SECTION 2: Data Refinement

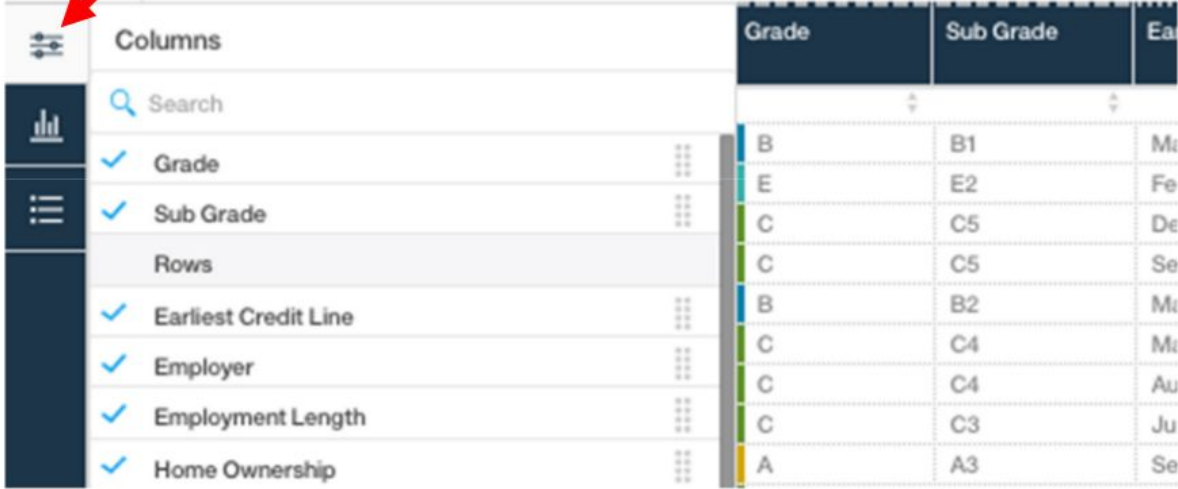1. click on the "..." ellipses on the bottom right of the uploaded dataset
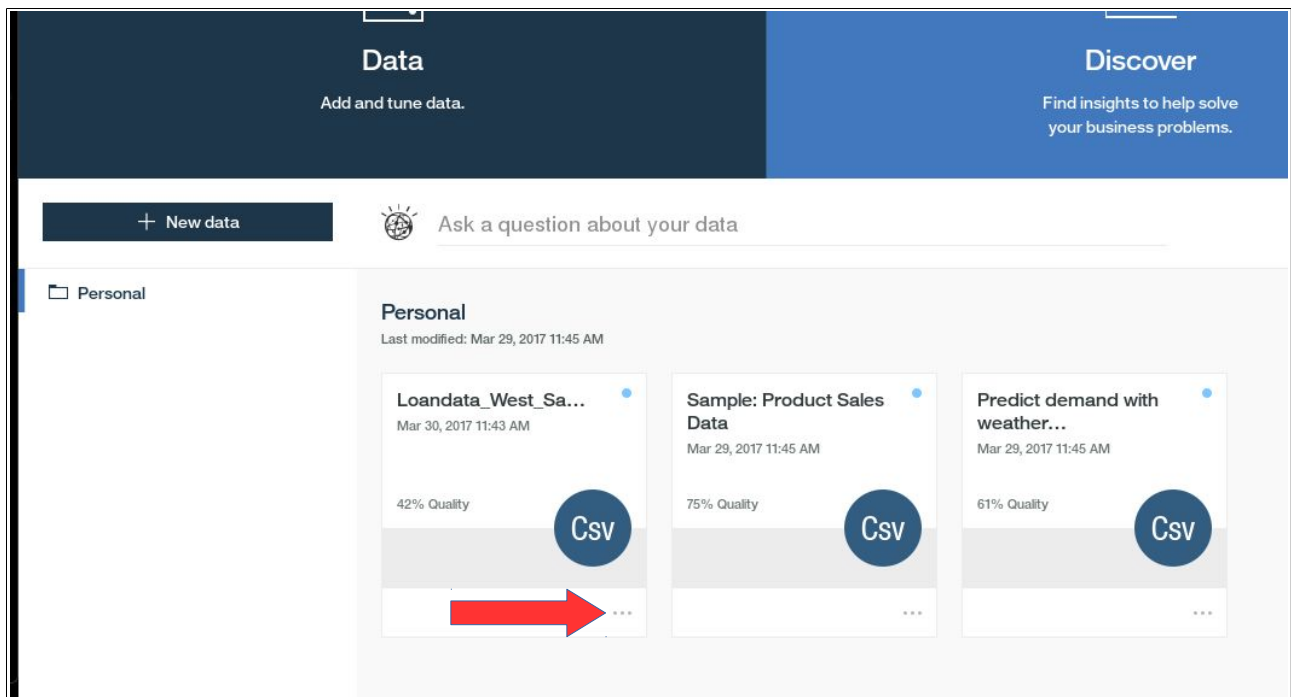
3. Click on "Refine"
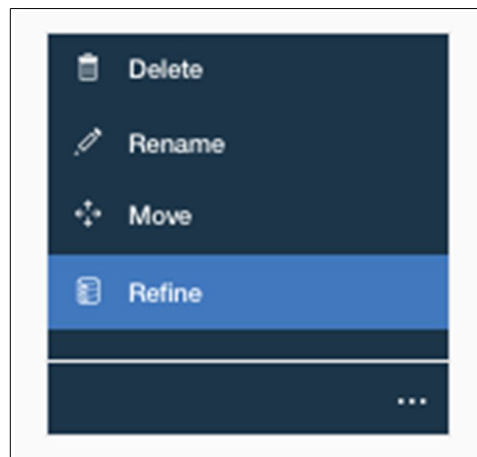
4. In Refine, we will make the custom data groups

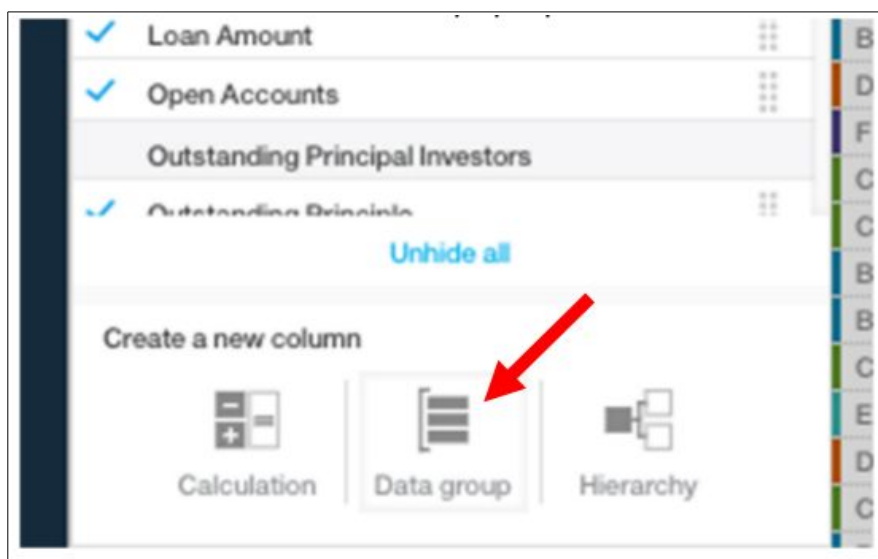5. Click on the first icon on the left navbar called "Actions"

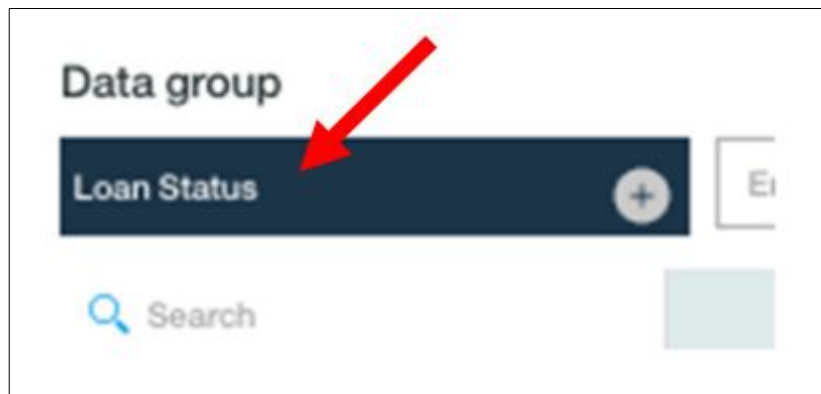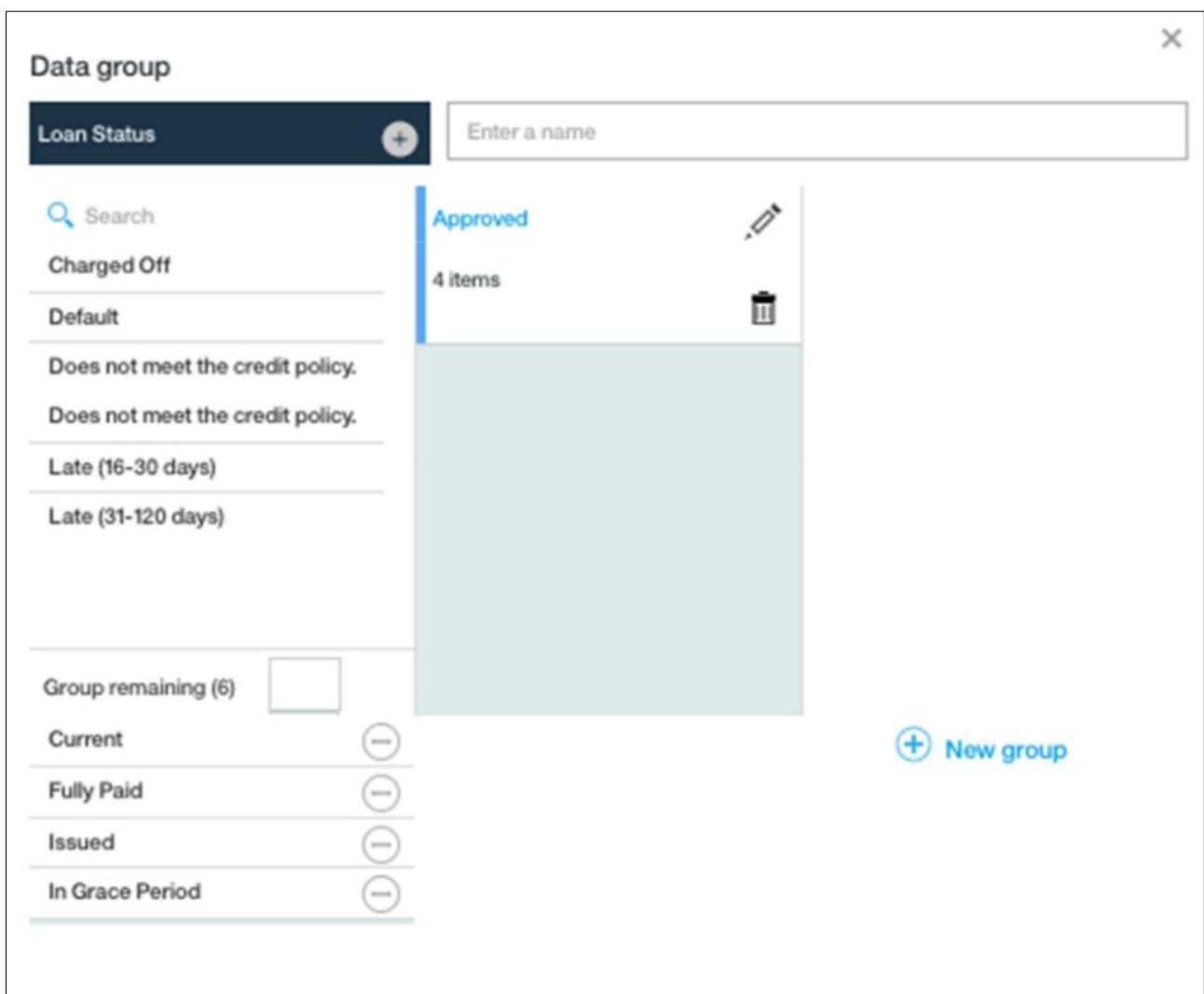6. At the bottom of the new pop-up window, click on "Data Group"

7. Choose "Loan Status" as the column to create the new data groups on

8. We will create 2 Data groups for "Approved" & "Rejected" loans.

9. Select: "Current","Fully Paid","Issued","In Grace Period" & click on (+) New Group. Label this new group as "Approved" ("Issued" may not be available in this data set. You can ignore it)

10. Click on "Group Remaining" (use the slider to enable it) and call it "Declined". Enter a name

for the new column. Name it: "Verdict" (top of screen). Click on Done. The new column will be created and added to the end of the data set.
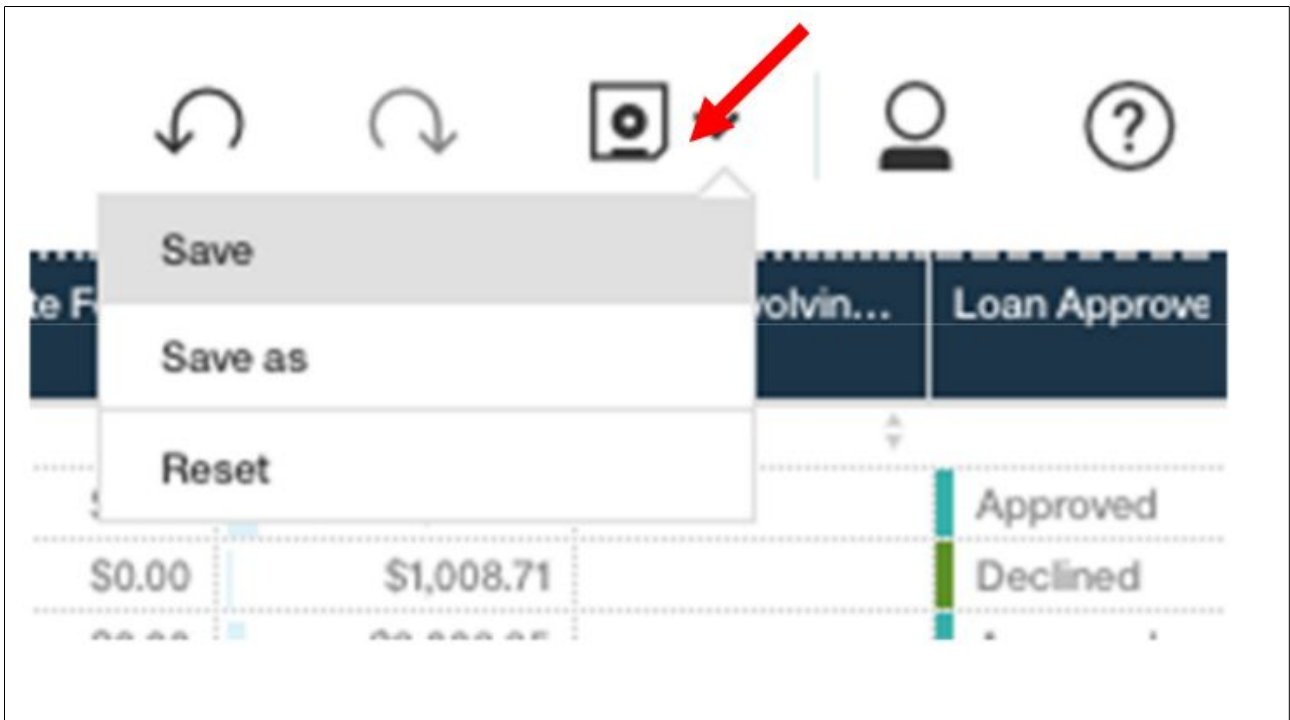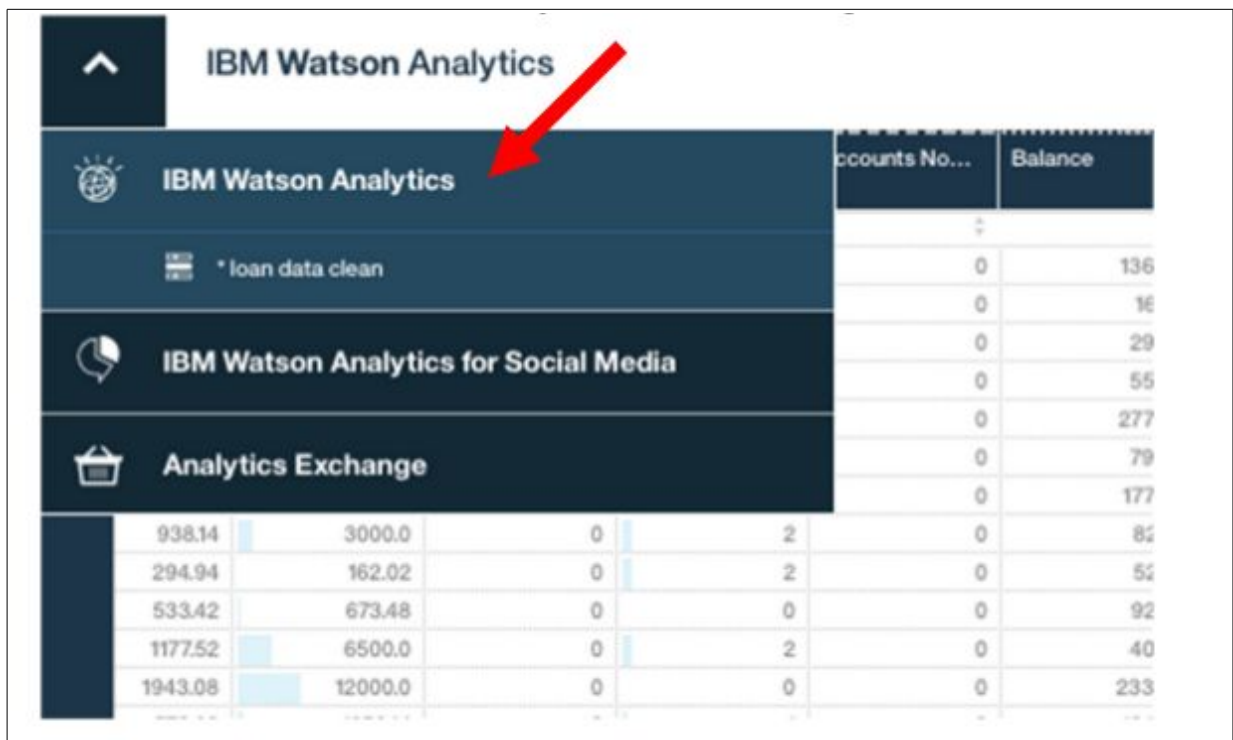

11. Save the refined data set. Click on the floppy disk icon on the top right navbar.

12. Navigate back to the Homepage. Click on the navigation icon on the top left navbar. Click on the Watson Analytics banner to get back to the home page.

SECTION 3: Data Exploration

1. Click on the saved data set tile. This will open the discovery zone. Here you can ask questions in natural language, access the auto generated insights, or create your own custom visuals from scratch.
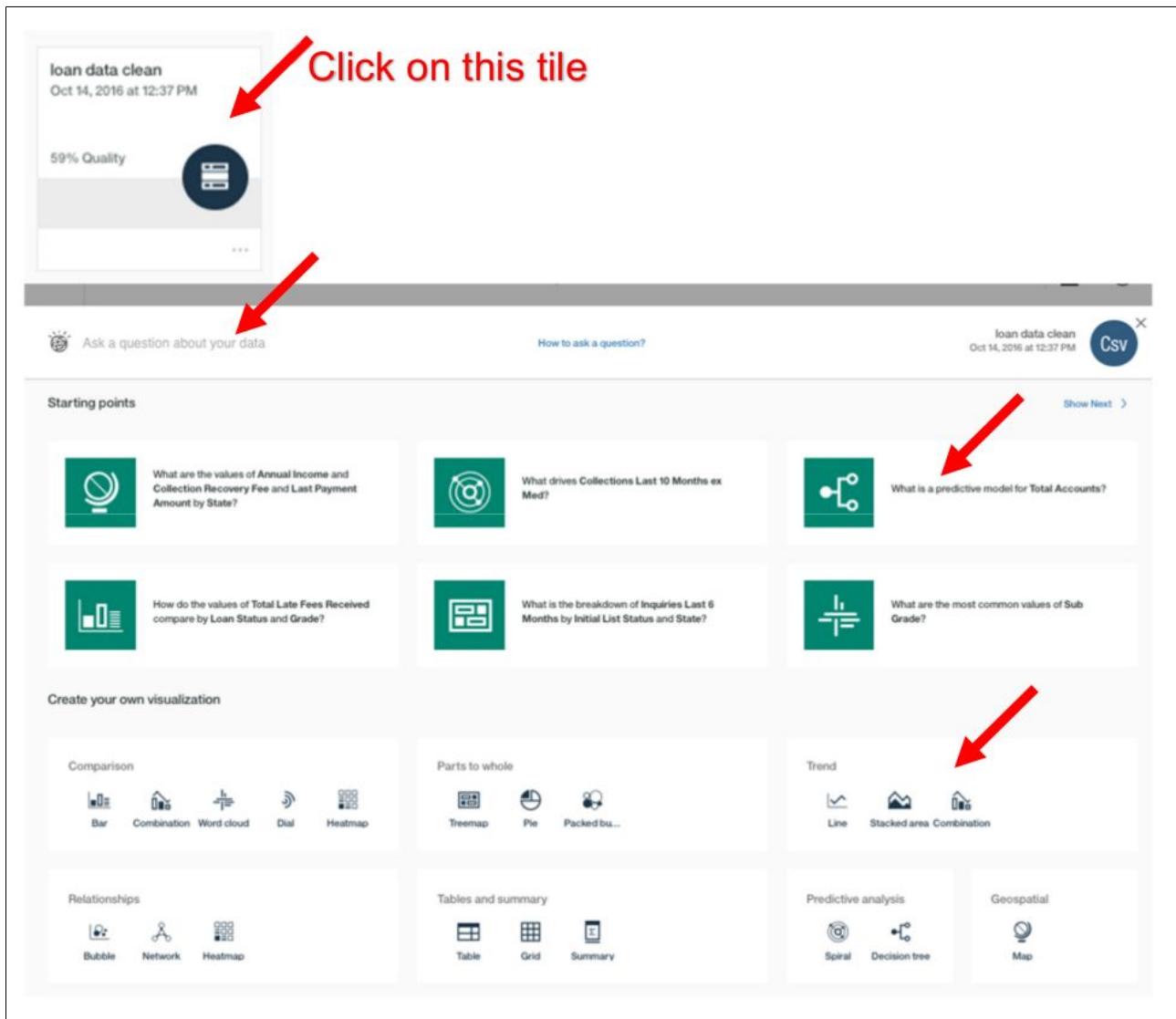
2. My first question is how many loan applications do we have for each state. In the natural language text bar, type in: "What is the number of rows in each state". Click on the suggestion showed below.

3. It looks like California, Texas and New York have the most amount of applications (the data set used in this lab only covers a sample of the western region of the US. Consequently, you will not see the whole map of the USA with all the states).

What is the number of **Rows** ⊗ by **State** ⊗ ?
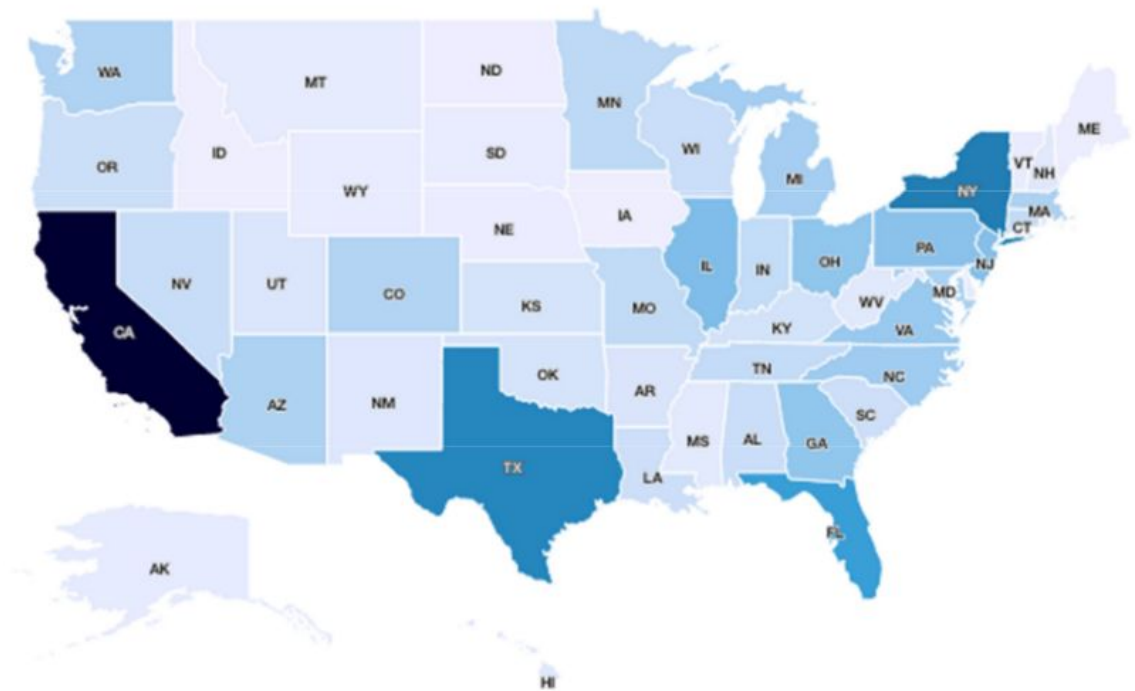
If you are interested in pursuing the data exploration in-depth, please check out the Watson Data Platform tutorial at https://github.com/ibmdataworks/datafirst