|  |
| --- |
| **Experiment No. 9**  **Title: Case study: Big data platform / analytics as business need)** |

**Batch: B1 Roll No.: 1824007** **Experiment No.:9**

**Title: Case study**

# Recommender: Movie recommendation

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Resources needed:** Microsoft Azure Machine Learning Studio

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Link to azure ml studio project:** https://gallery.cortanaintelligence.com/Experiment/Movie-recommendation-5

**Describe the following points with respect to the business under consideration,**

1. **Problem faced by the business**

# Recommender: Movie recommender

The main aim of a recommendation system is to recommend one or more items to users of the system. Examples of an item could be a movie, restaurant, book, or song. A user could be a person, group of persons, or other entity with item preferences.

1. **Approach/ Methodology followed by the business**

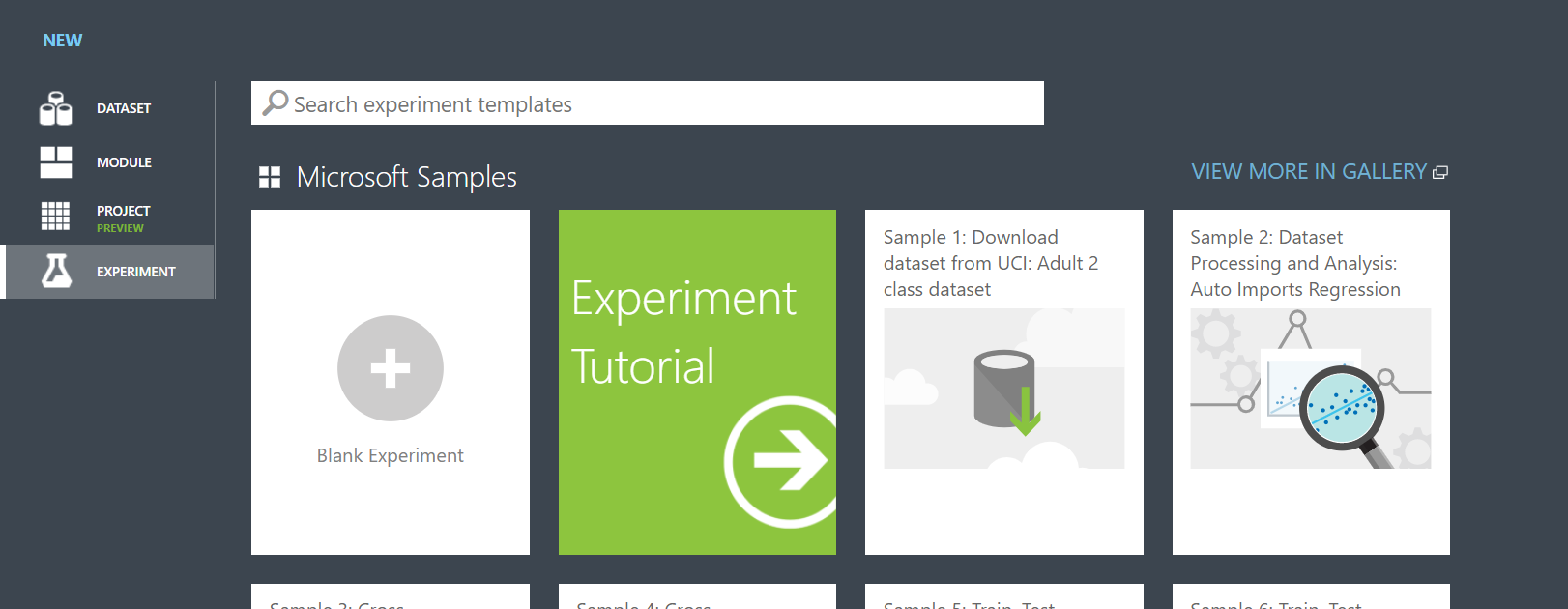
There are two principal approaches to recommender systems.

* The first is the content-based approach, which makes use of features for both users and items. Users may be described by properties such as age and gender, and items may be described by properties such as author and manufacturer. Typical examples of content-based recommendation systems can be found on social matchmaking sites.
* The second approach is collaborative filtering, which uses only identifiers of the users and the items and obtains implicit information about these entities from a (sparse) matrix of ratings given by the users to the items. We can learn about a user from the items they have rated and from other users who have rated the same items.

The Matchbox recommender combines these approaches, using collaborative filtering with a content-based approach. It is therefore considered a hybrid recommender.

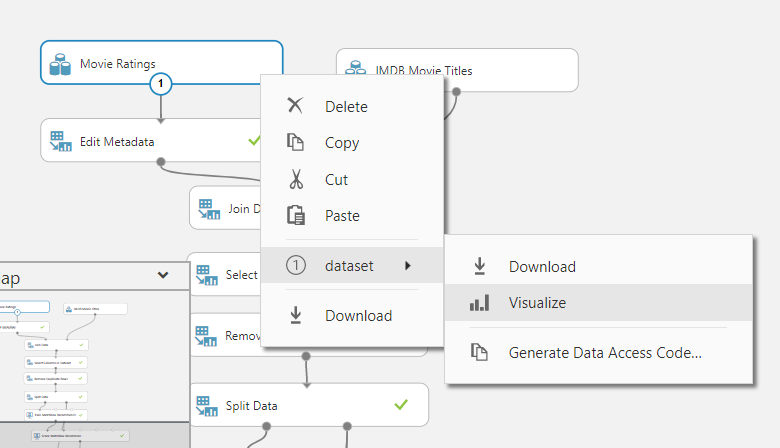
**Skill-sets, infrastructure and other impact on the business during implementation**

Open the Azure Machine Learning Studio and create a new blank experiment.

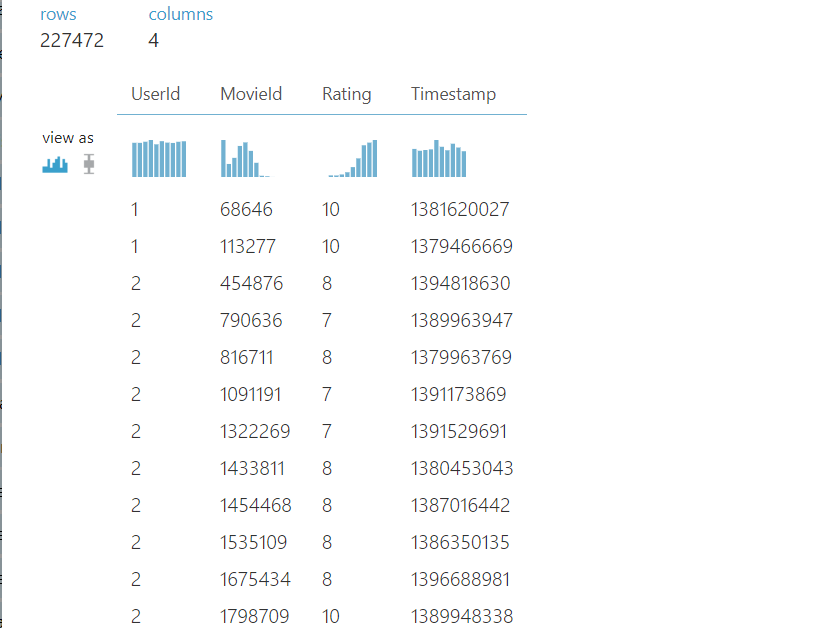
****

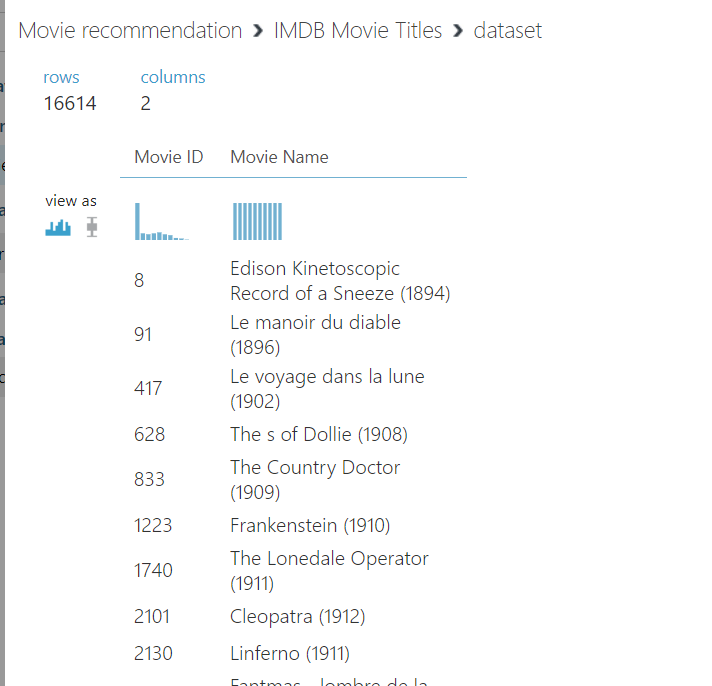
Then find in the left column 2 datasets, in the Datasets section, the IMDB Movie Titles and Movie Ratings and drag and drop them in the experiment area.

To see the dataset contents right click on the dataset and click Visualize.



## Data:

The training data is approximately 225,000 ratings for 15,742 movies by 26,770 users

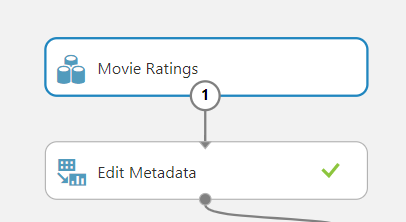
To this data, we added a file containing movie names extracted from IMDB, joined on the movie identifier from the ratings data. 

Next add the Edit Metadata module and connect it to the first dataset as in the following image. To connect the dataset put your mouse to the dot you want to connect and click and drag to the destination dot in the Edit Metadata Module.

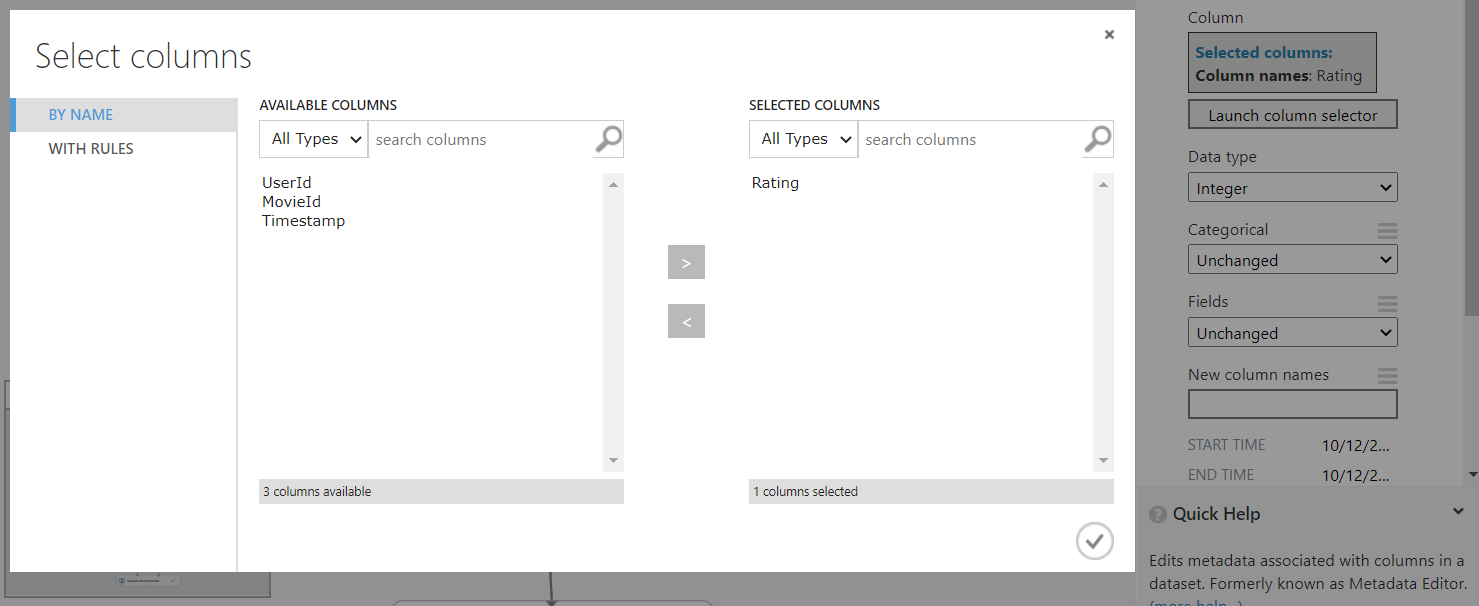
This module edits metadata associated with columns in a dataset. Typical metadata changes might include:

* Treating Boolean or numeric columns as categorical values
* Indicating which column contains the class label, or the values you want to categorize or predict
* Marking columns as features
* Changing date/time values to a numeric value, or vice versa
* Renaming columns

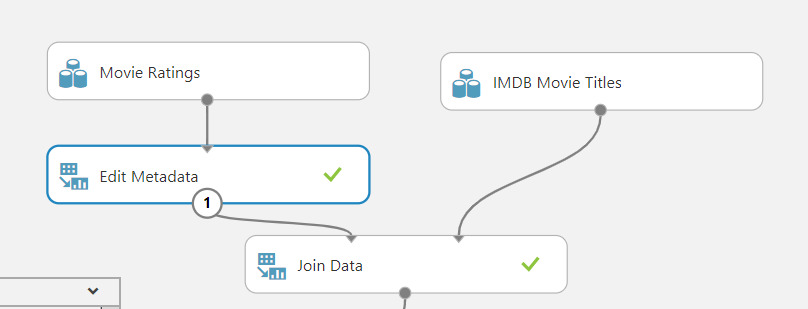
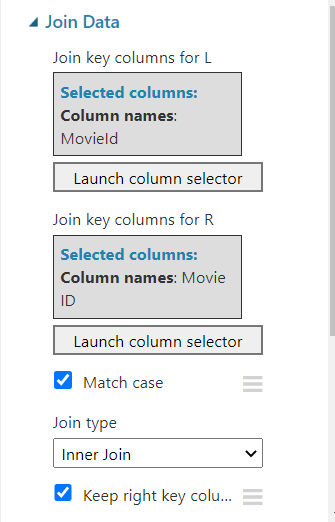
Use Edit Metadata any time you need to modify the definition of a column, typically to meet requirements for a downstream module. For example, some modules can work only with specific data types, or require flags on the columns, such as IsFeature or IsCategorical. After performing the required operation, you can reset the metadata to its original state. Here we are going to use the Edit Metadata to convert the Rating Column into an integer, so it can be used by the machine learning algorithm.



Click on the module and then click on the Launch column selector button in the pane that loads in the right. Choose Rating and close the window. Then in the pane on the right on the first dropdown choose Data type→ Integer



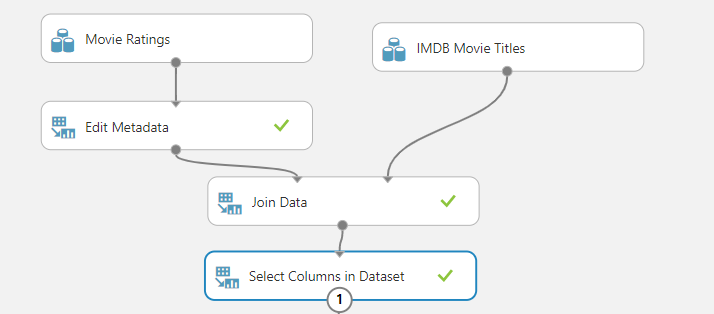
Add the Join Data module. The Join Data module joins two datasets. In the right pane choose MovieId from Movie Ratings and Movie ID from IMDB Movie Titles, so the results can show the Title instead of the Movie Id.

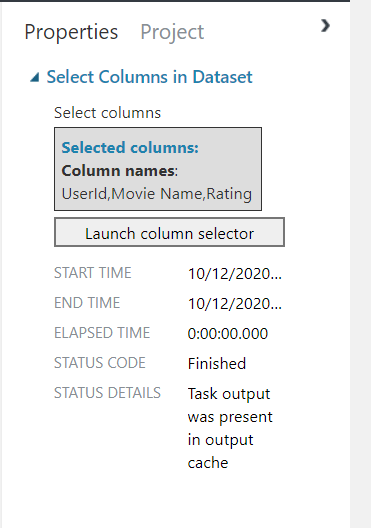
 

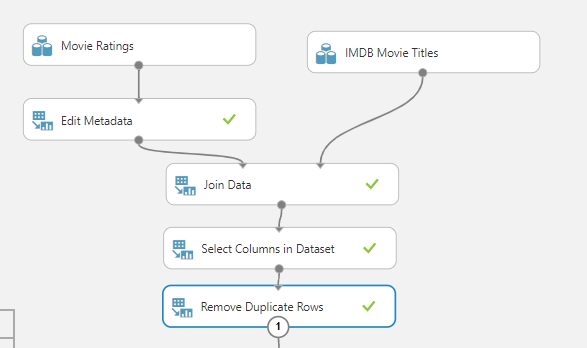
It is very important that the input data used for training contain the right type of data in the correct format:

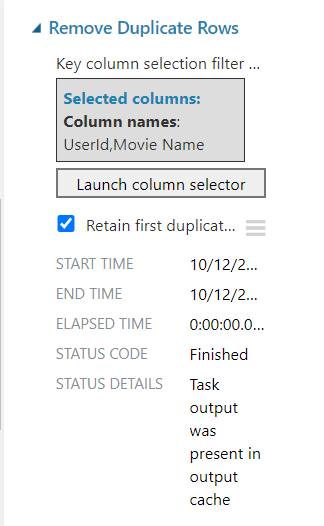
* The first column must contain user identifiers.
* The second column must contain item identifiers.
* The third column contains the rating for the user-item pair. Rating values must be either numeric or categorical.

During training, the rating values cannot all be the same. Moreover, if numeric, the difference between the minimum and the maximum rating values must be less than 100, and ideally not greater than 20.

Add the Select Columns in Dataset Module to select the proper columns for training. Select the columns UserId, Movie Name, Rating. 

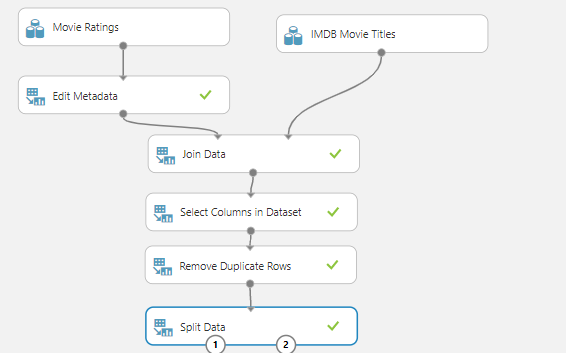


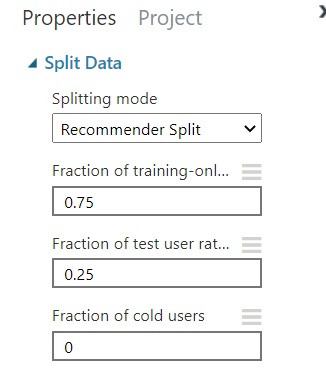
Then add the Remove Duplicate Rows module to remove duplicates, as not all items must be the same, meaning a user cannot have more than one rating for the same Movie. Use the combination of UserId and Movie Name in the column selector as shown below. 

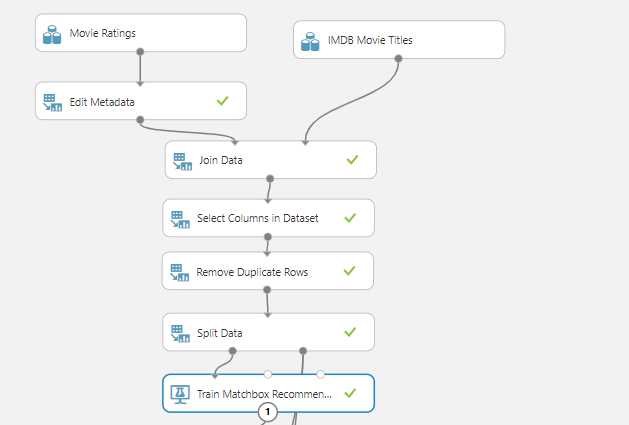
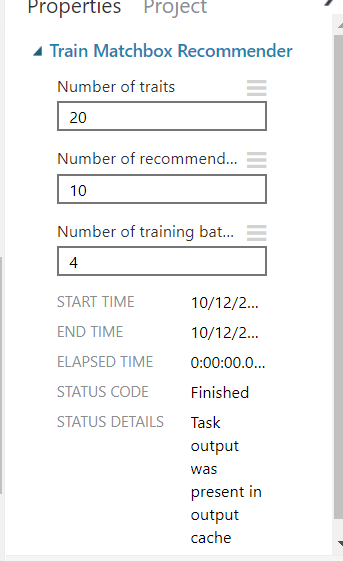


Next we need to divide source data into training and testing datasets. We use the training dataset to train our recommender module and then we use the testing dataset to test and score the results. To accomplice this we are going to use the Split Data Module. This module is particularly useful when you need to separate data into training and testing sets. You can customize the way that data is divided as well. Some options support randomization of data; others are tailored for a certain data type or model type.

Add the Split Data module and use the **Recommender Split** option.

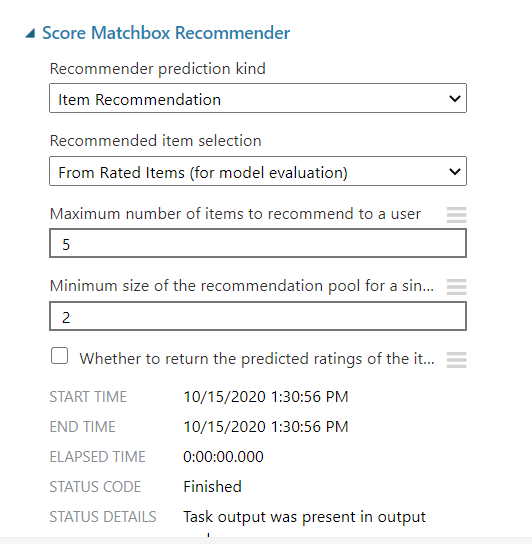




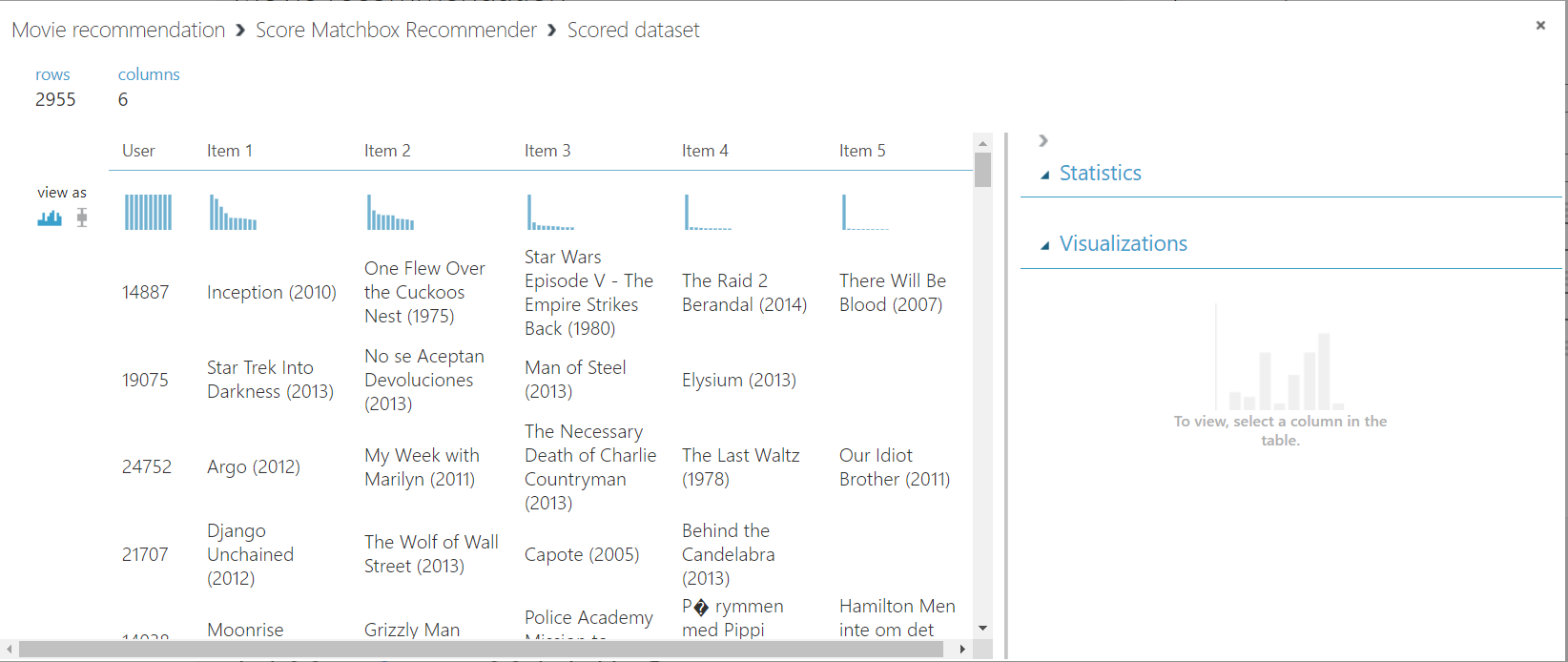
Then the next step is to add the Train Matchbox Recommender, to train our recommender model. 

Add the Score Matchbox Recommender using the following options

* Item Recommendation
* From Rated Items

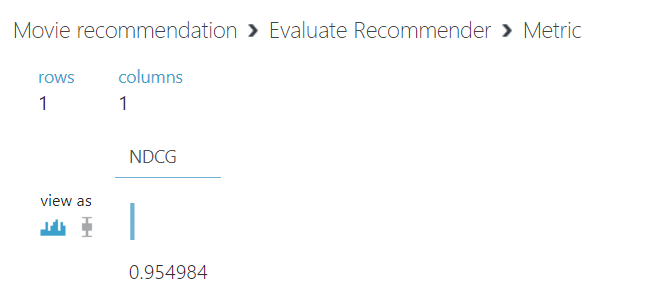


**Results** of score matchbox recommender which displays upto 5 recommendation, item1 as highly recommended and item 5 as low recommendation for the user.



Finally add Evaluator recommender and left click to visualize data (metric-> visualize) to see the accuracy of the Module.

**95% accuracy**.



**Questions:**

**Discuss the tangible and intangible benefits the business has observed after the implementation.**

1. **Budget Management:**

One of the major tangible benefits of project management is budgetary savings. Project managers control budgets and make decisions about how best to allocate resources in the process of working toward a project's objectives. In the end, the difference between coming in under budget or having an overrun is a function of workforce efficiency and project management. The money that project managers save their businesses increases the project's return on investment and remains available for future projects.

1. **Managing Risk:**

Rick management is much more difficult to quantify than cost savings. Project managers can anticipate risk and guide a project to best avoid it. While there are some financial metrics for placing a tangible value on risk, project management also has the intangible benefit of taking on risks only when they are necessary or worthwhile in the first place. A skilled project manager can lead a project team toward its objectives but remain flexible to react to changing risk factors.

1. **Meeting Deadlines:**

Project management is also concerned with managing time and meeting deadlines. Completing a project on time can be as important as completing it on budget. Projects that take too long to complete cost money in terms of overtime wages, lost productivity and time spent modifying schedules and timelines. Project management's impact on meeting deadline is a tangible benefit when the costs of late completion are known. For example, if a business spends $100,000 each day operating a factory to meet a production quota, the tangible benefit of project management that competes the project on time is $500,000 for every week of estimated late delivery.

1. **Teamwork:**

Another intangible benefit of project management is its effect on teamwork within an organization. One of the key areas project management deals with is the allocation of human resources. Successful project management gets the most out of each worker and fosters an environment of cooperation and mutual responsibility that can remain long after the project is completed. The intangible benefit of a productive, collaborative workforce is a fundamental part of a successful business's organizational culture despite the impossibility of assigning a dollar value to it.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Outcomes:** Realise adequate perspectives of big data analytics in various applications

**Conclusion: (Conclusion to be based on the objectives and outcomes achieved)**

# In this given experiment, I have implemented movie recommendation system.

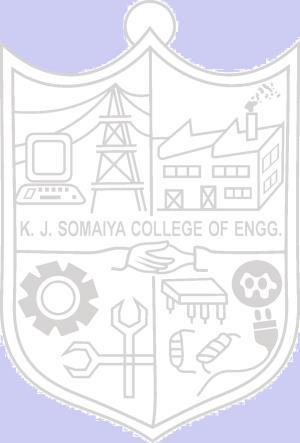
This experiment demonstrated the use of the Matchbox recommender modules to train a movie recommendation engine. I have understood the tangible and intangible benefits the business after the implementation of the model.

**Grade: AA / AB / BB / BC / CC / CD /DD**

**Signature of faculty in-charge with date**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**References:**



**Books/ Journals/ Websites:**