Company Name: FirstEigen

Name: Aishwarya Patil

Joining date: 27th December, 2021

College: AISSMS COE, Pune

Year & Department: Third Year, Computer Dept.

**Suggestions for Databuck VS Collibra**

**Databuck Functionalities:-**

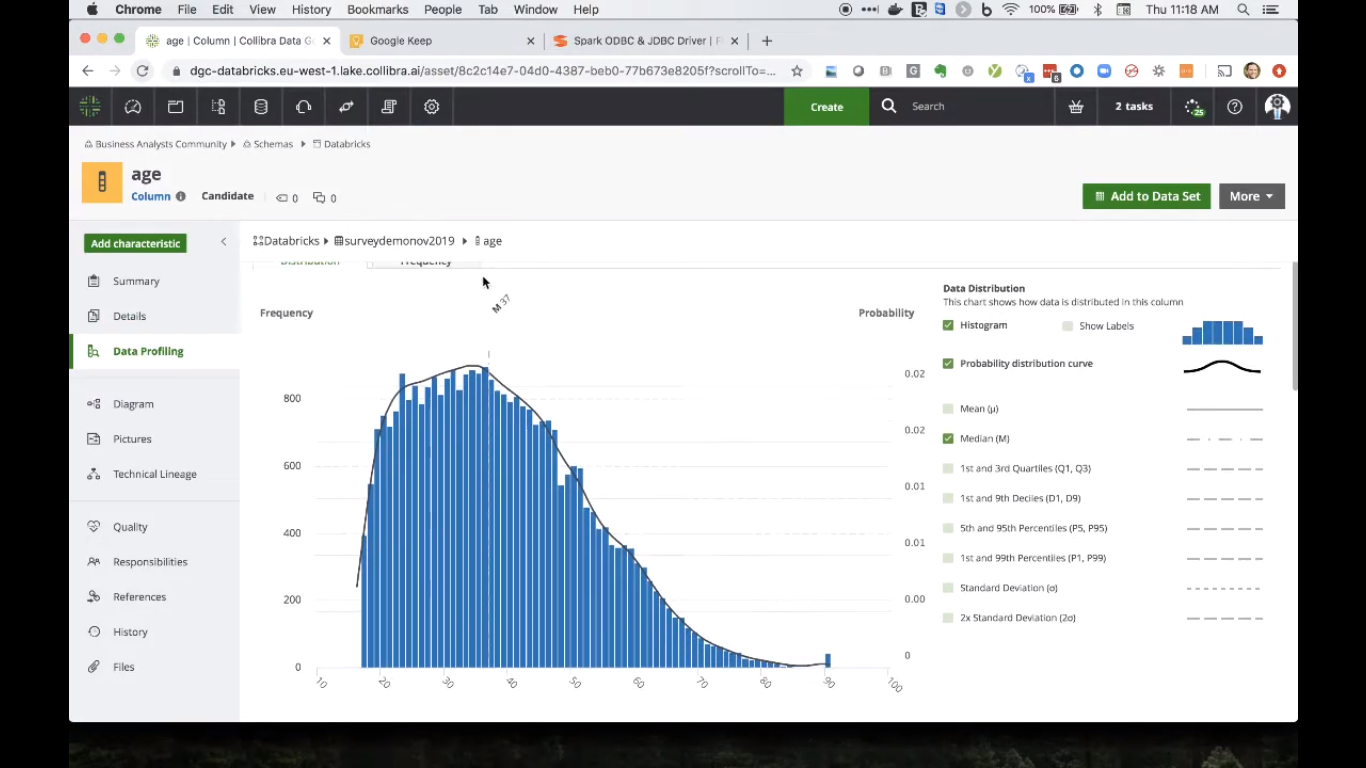
With just a few clicks and the power of AI/ML, 7 key data monitoring functionalities can be operationalized on AWS S3.

It increases the scalability of continuously monitoring data to thousands of incoming data sets. It eliminates months of labour by using Machine Learning.

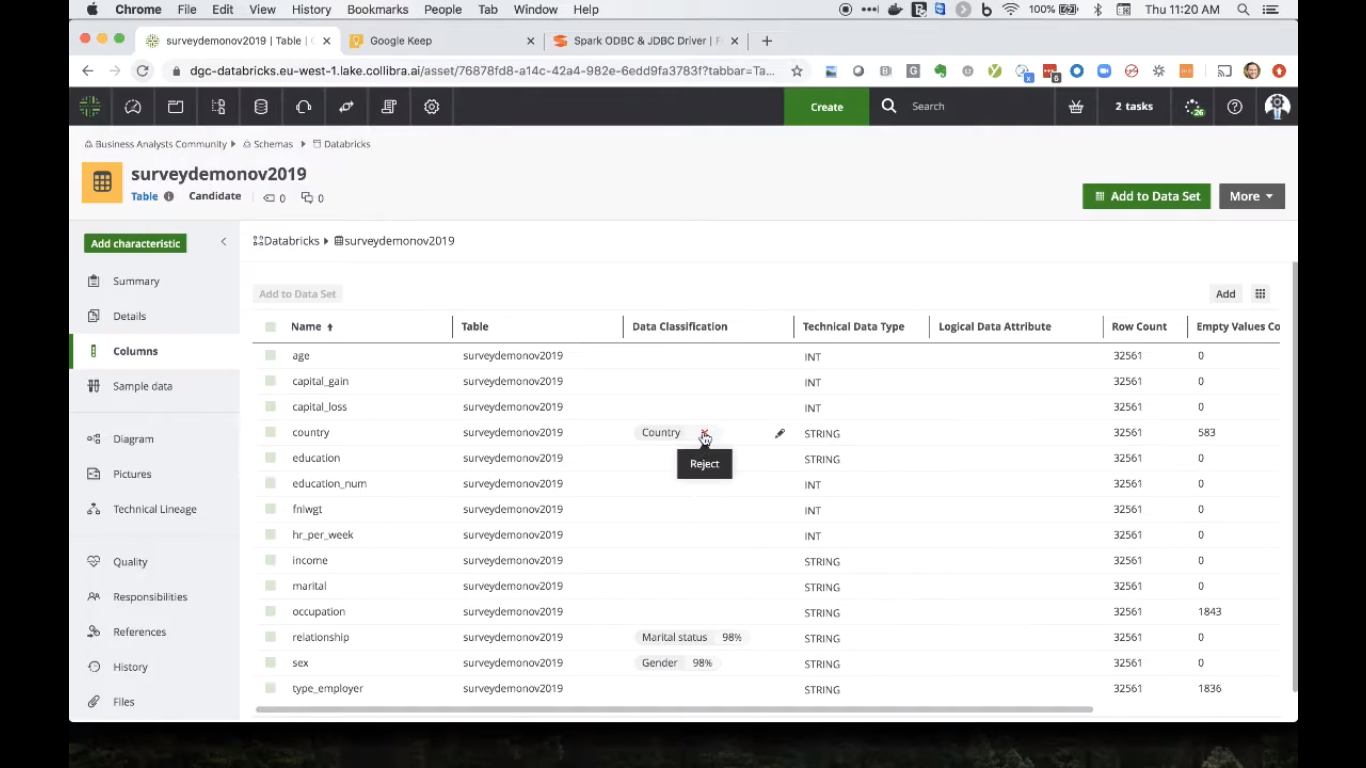
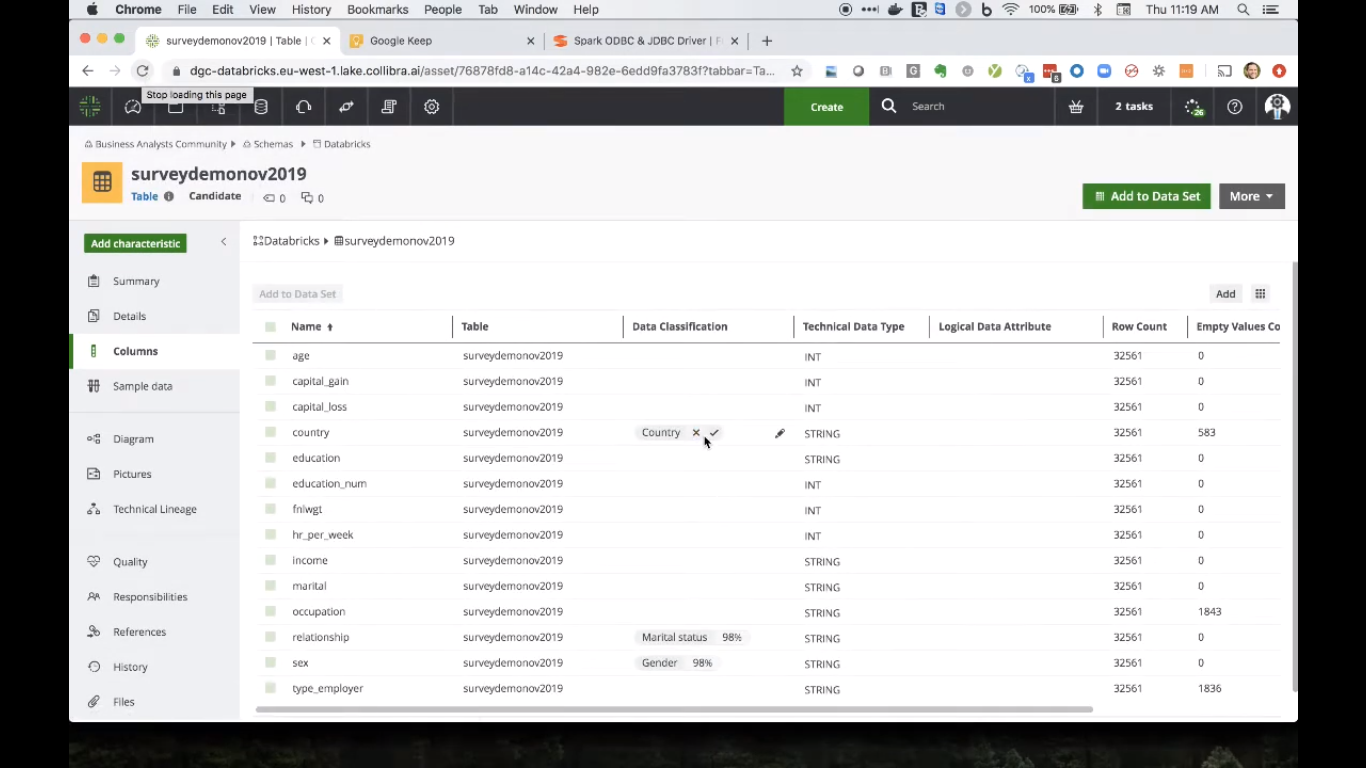
* Data Quality Health Score of data set
* Freshness of data set
* Nulls
* Duplicates
* Format deviations
* Record Count
* Data Profile

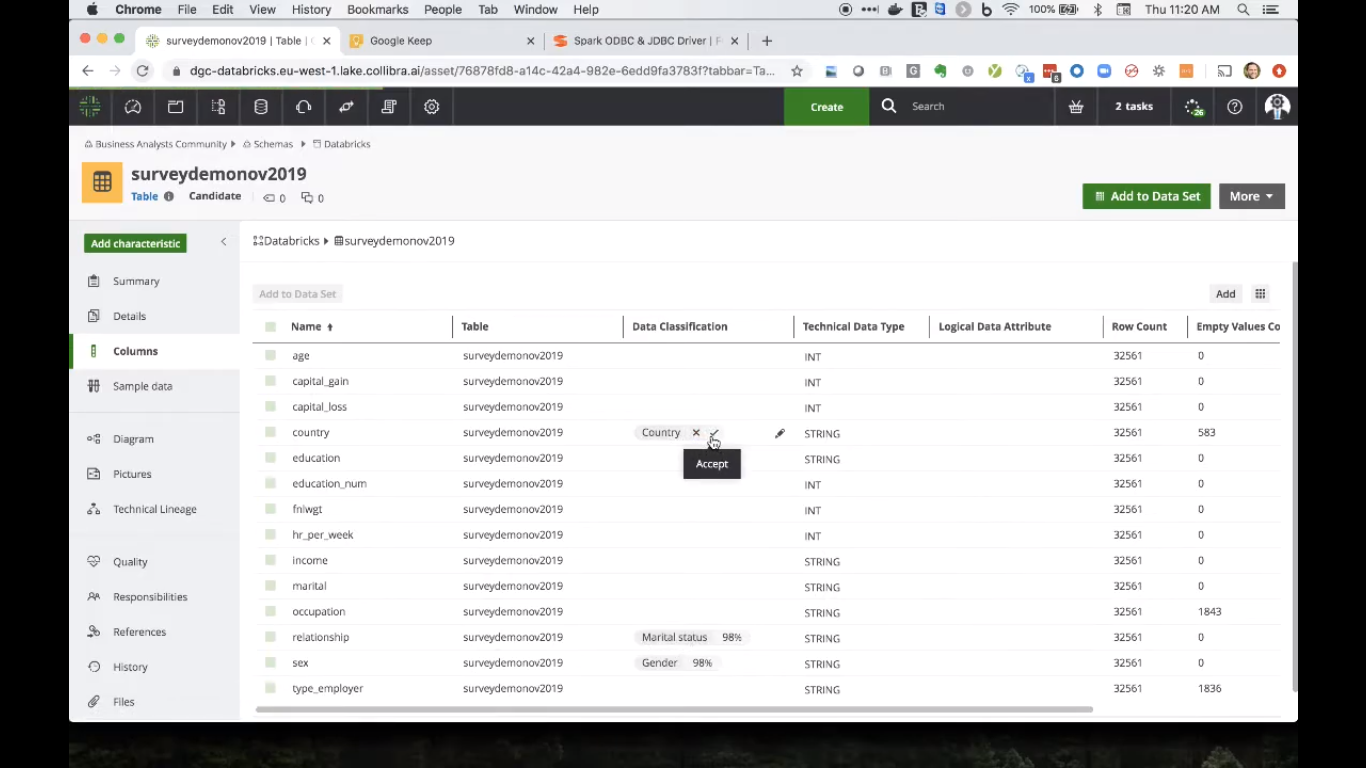
**Suggestions for databuck:-**

1. Databuck can be improved to build a profile of each column with a summary and most detailed profiling using histogram graph pattern on the sample data in the columns as shown in the image.
2. Based on available sample data it builds distribution and frequency charts and has different kinds of medians, percentiles, etc.



1. Once we classify the data manually or automated then Databuck can check the correctness of data. Here the user can give feedback whether he/she agrees or disagrees with the results. This action of accepting/rejecting gets sent back to the Databuck ML model and it learns and updates itself accordingly. This same functionality takes place in collibra which can be implemented in Databuck.





1. We can search for any metadata or column name, etc. So same like that we can add some conceptual elements to represent the technical(abbreviated) column names because users won't know what the abbreviation indicates and therefore save the corresponding representation in the glossary search for further use in Databuck too.

