MSA 8040 Data Management for Analytics:

Final Project Report:

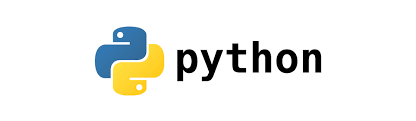
Data Analysis with Estimize.com

Shreyashi Mukhopadhyay (Panther id: 002592906)

**Tools used: Selenium, MySQL, Python**

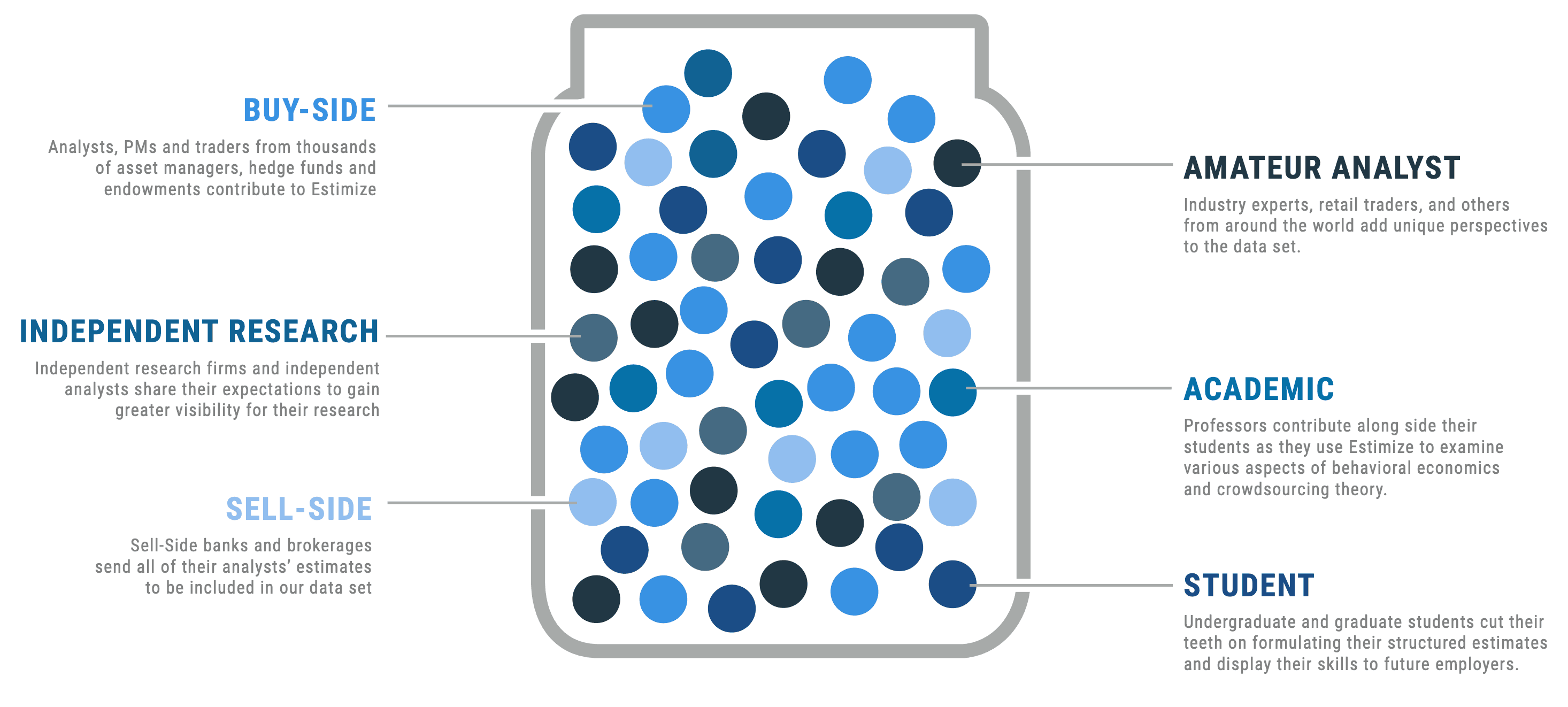
Logo, company name

Description automatically generated

Icon

Description automatically generated

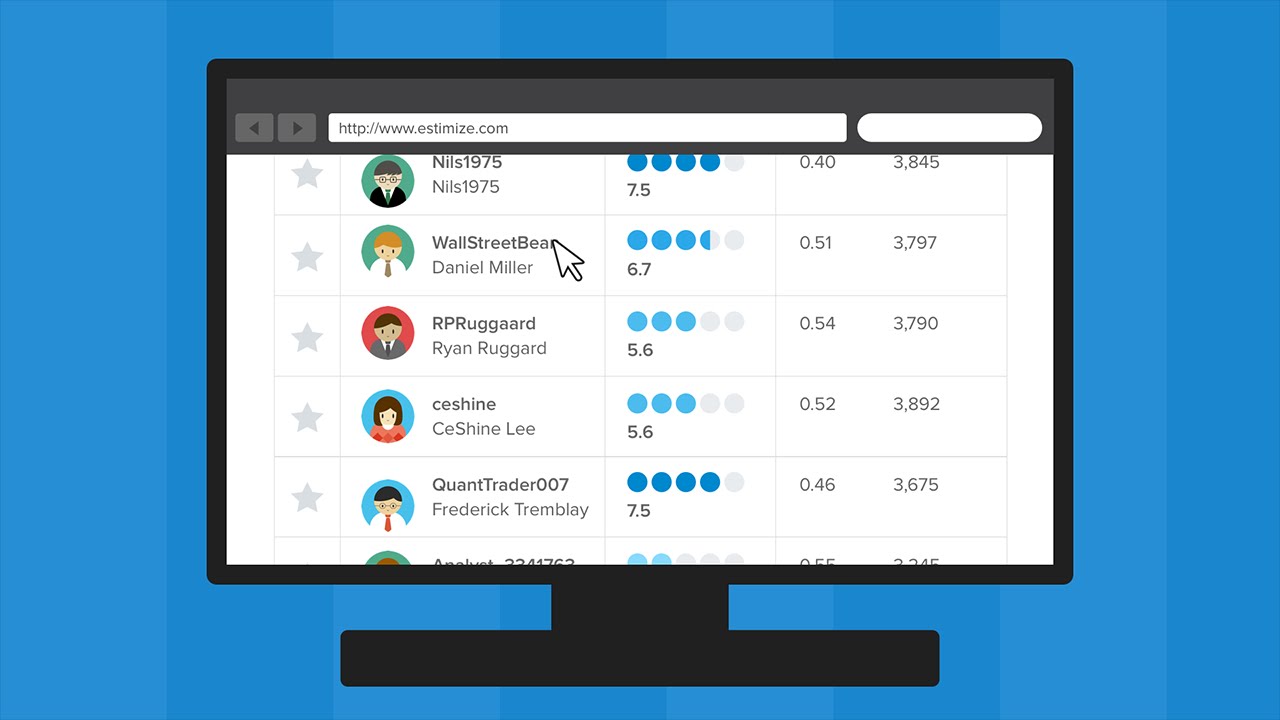
**About Estimize.com**

Estimize.com is a free and open financial estimates platform where hedge fund, independent, and sell-side analysts, along with private investors, industry experts and students contribute their EPS and Revenue estimates for public companies. By sourcing estimates from a diverse community of individuals, Estimize provides both a more accurate and more representative view of expectations compared to sell side only data sets.

Launched in 2011, Estimize is an open financial estimates platform designed to collect forward looking financial estimates from independent, buy-side, and sell-side analysts, along with those of private investors and academics.

By sourcing estimates from a diverse community of individuals, Estimize provides a more accurate, more timely, and more representative view of expectations compared to sell side only data sets which suffer from several severe biases.

Currently, 96,211 analysts contribute to Estimize, resulting in coverage on over 2,200 stocks and 80 economic indicators each quarter. The Estimize consensus has proven more accurate than comparable sell side data sets over 70% of the time and by 15% on average.

**How does Estimize.com work?**

Contributors to the Estimize platform receive free access to view their data in return for their honest pseudonymous contributions. Estimize manages the honesty and quality of contributions via several machine learning algorithms and statistical methods, along with a human layer of review. Estimize scores and ranks analysts, providing them with an easy way to store, benchmark, and measure their own accuracy.

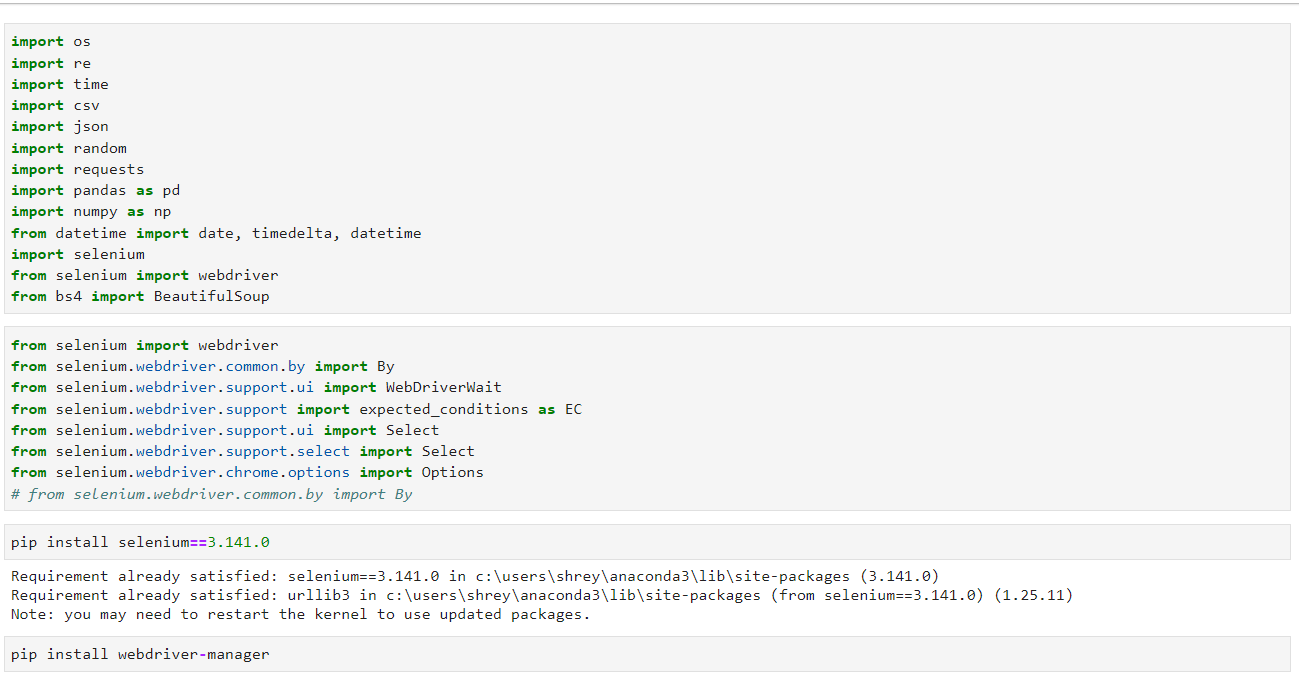
Their data is available to purchase via the Estimize web platform, their API, and FTP and does not require you to contribute. Their clients include institutional managers running fully systematic quantitative strategies, fundamental managers running quanta mental, long/short, and long only strategies, sell side market making and research desks, as well as macro investors and vol traders. Estimize represents the market’s true consensus, with their data regularly referenced in notable financial media sources such as Bloomberg, The Wall Street Journal, CNN Money, The Street, Forbes, Barron’s, Investor’s Business Daily and Business Week, amongst others.

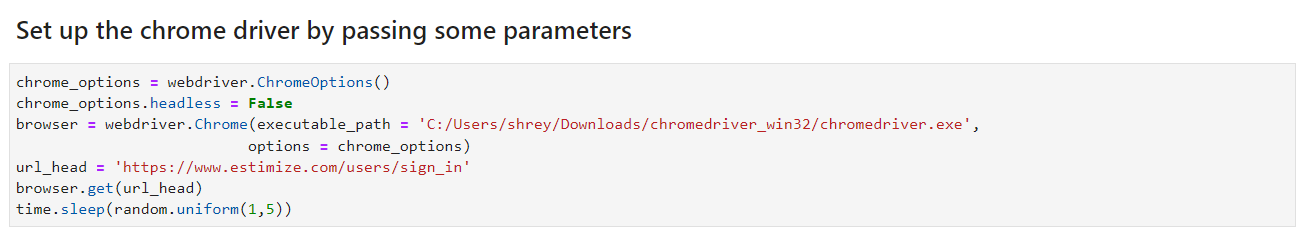
## Problem Statement:

* Scrape the data for 50 tickers from Estimize.com for at least one year – 4 quarters.
* Build datasets to store the scraped EPS, Company and Analyst information.
* Build a database using the datasets generated above that supports easy query for extracting the correct information.

## Problem solution & procedure:

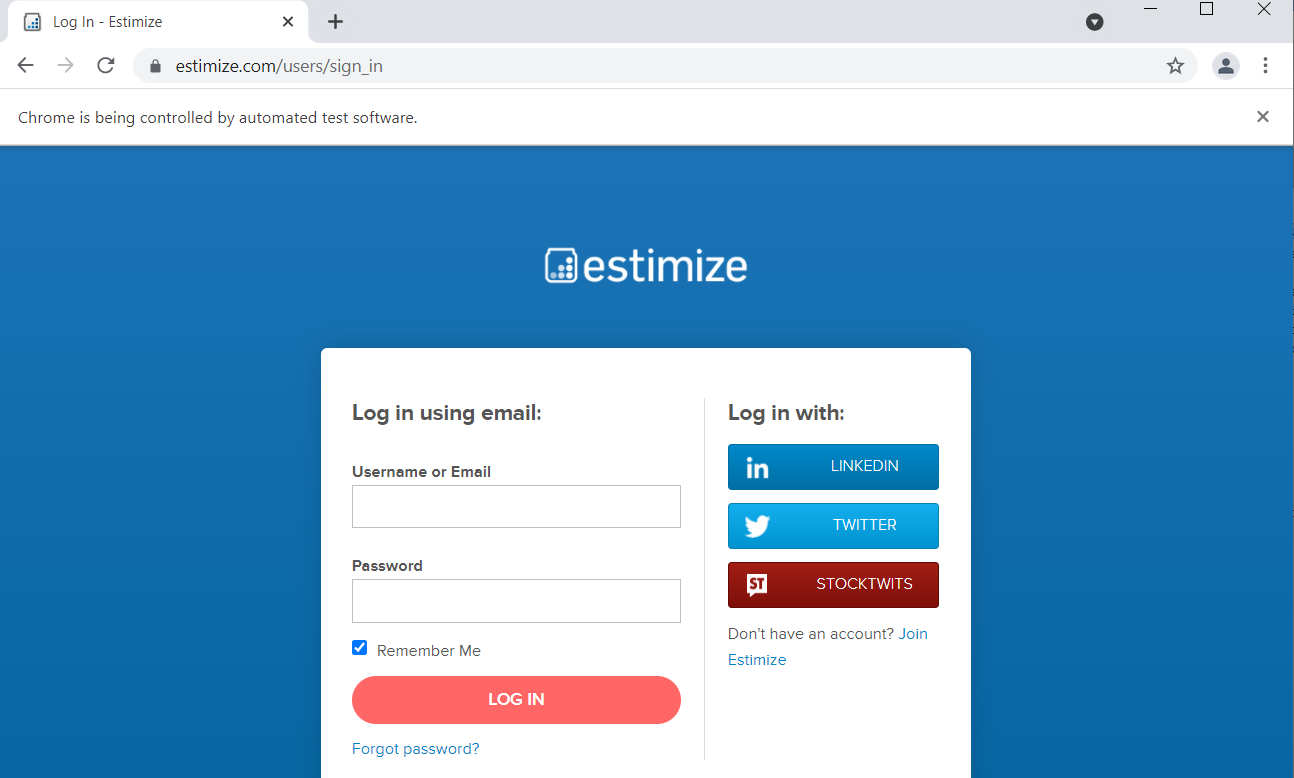
* Data on Estimize.com is unstructured data which is not readily downloadable or usable.
* We will scrape the data from Estimize.com using **Python and** **Selenium.**
* We will then store the scraped data into a json or csv format.
* Finally, we will import the stored data in **MYSQL** workbench and query the data as per further requirements.

**1. Setting up the environment in Jupyter Notebook:**

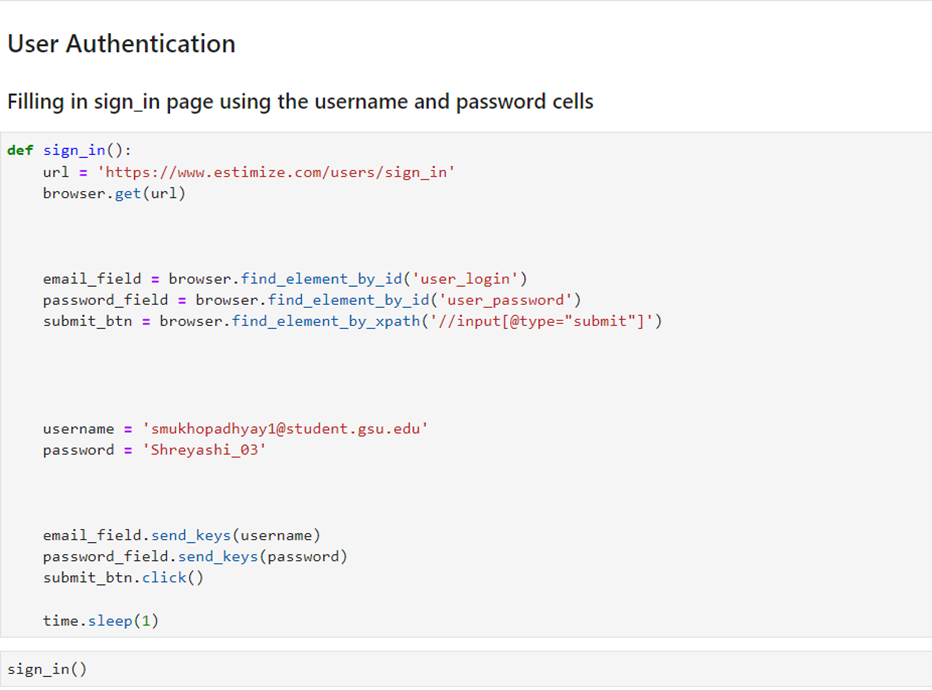
**2. Setting up the chrome driver and defining the browser object**

**3. Chrome launches the Estimize user sign-in page in the browser:**

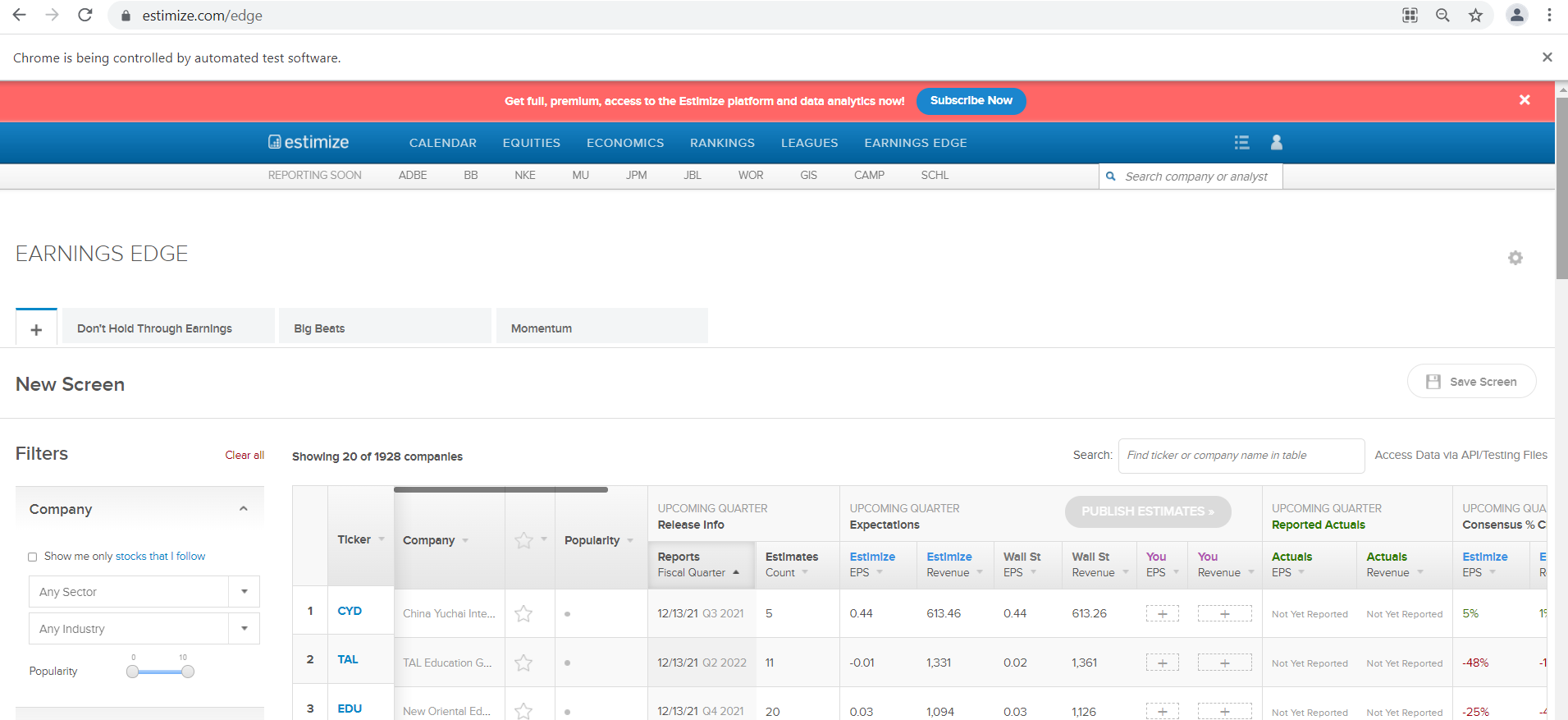
<https://www.estimize.com/users/sign_in>

****

**3. User authentication using credentials in the sign-in page**

****

**4. User page is displayed upon successful login.**

****

**5. Ticker and quarter selection.**

**6. Building the Company table:**

The following section of the webpage was scraped to generate the company data fields.

<https://www.estimize.com/aal/fq1-2020?metric_name=eps&chart=historical>



The company basic information, such as Ticker, company name, Sectors, Industries, number of followers, number of analysts.

i. ticker: AAL

ii. name: American Airlines Group Inc.

iii. Sectors: Industrials

iv. Industries: Airlines

v. number of followers: 618

vi. number of analysts: 589

**Python and Selenium code for generating data for the Company table:**

Graphical user interface, text, application

Description automatically generated



**Company.csv**

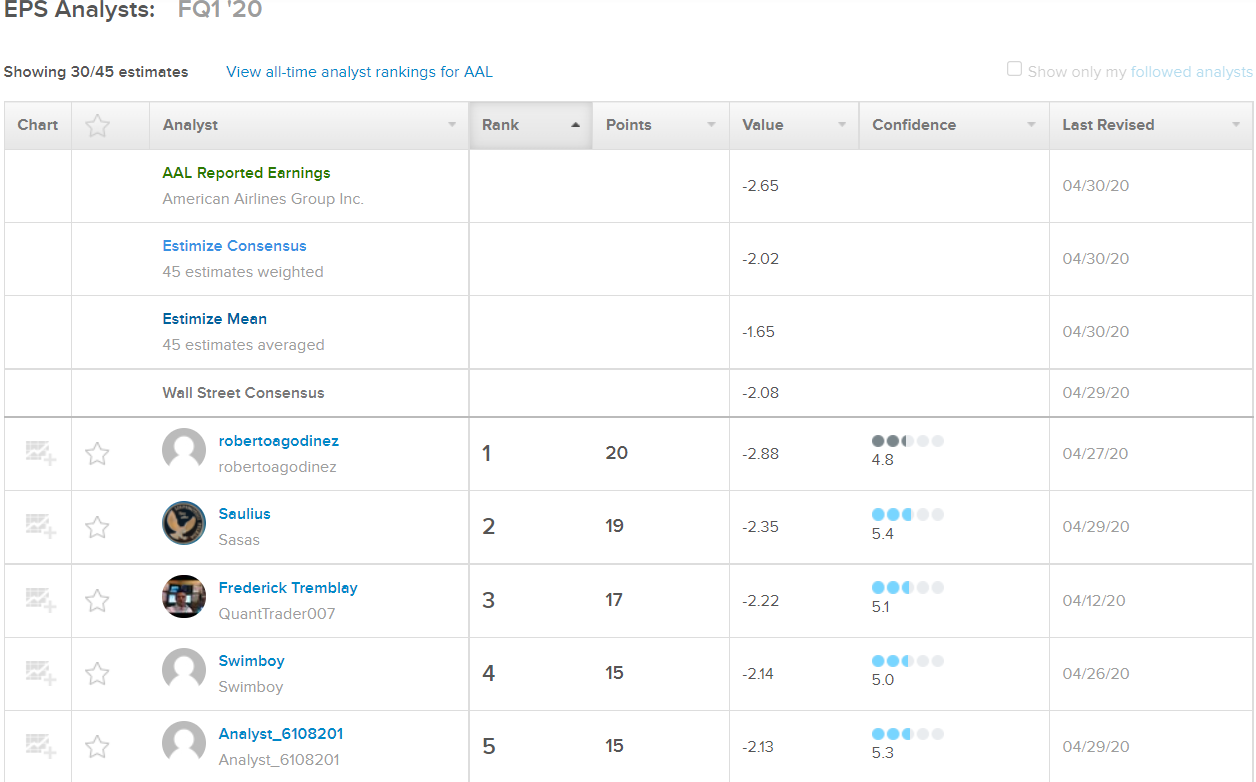
**Building the EPS table:**

This table contains the EPS information, including:

1. Reported Earnings
2. Estimize Consensus
3. Estimize Mean
4. Wall Street Consensus
5. EPS estimations of all available analysts

The following section of the webpage shown below was used to scrape the EPS Analyst data fields.

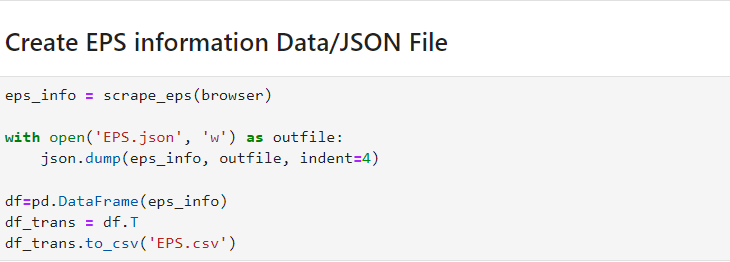
<https://www.estimize.com/aal/fq1-2020?metric_name=eps&chart=historical>



**Python and Selenium code for generating data for the EPS table:**

Graphical user interface, text, application

Description automatically generated



**Graphical user interface, application

Description automatically generatedEps.csv**

**Building the Analyst table:**

This table contains the information about each analyst, including Analyst name, Roles, Join date, Analyst Confidence score, number of estimates, Stocks Covered, pending estimates, Scored estimates.

i. name: Bill

ii. roles: Non Professional Financials Professional Services

iii. Join Date: Jul 2014

iv. Analyst Confidence Score: 8.4

v. error rate: 16.5%

vi. Accuracy Percentile: 59%

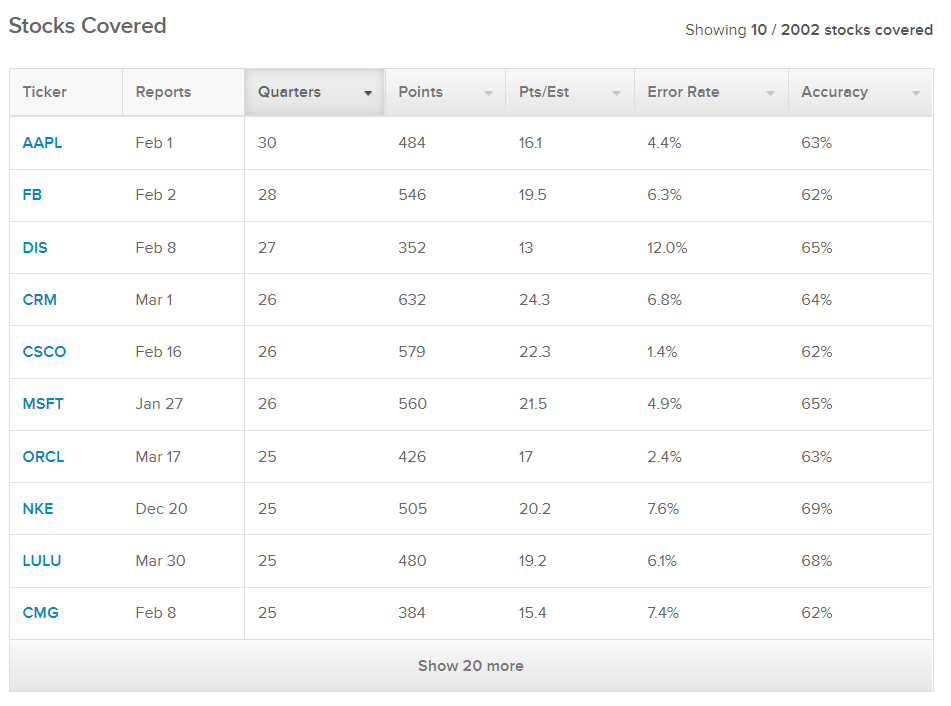
vii. points: 25,714

viii. points/Estimate: 13.6

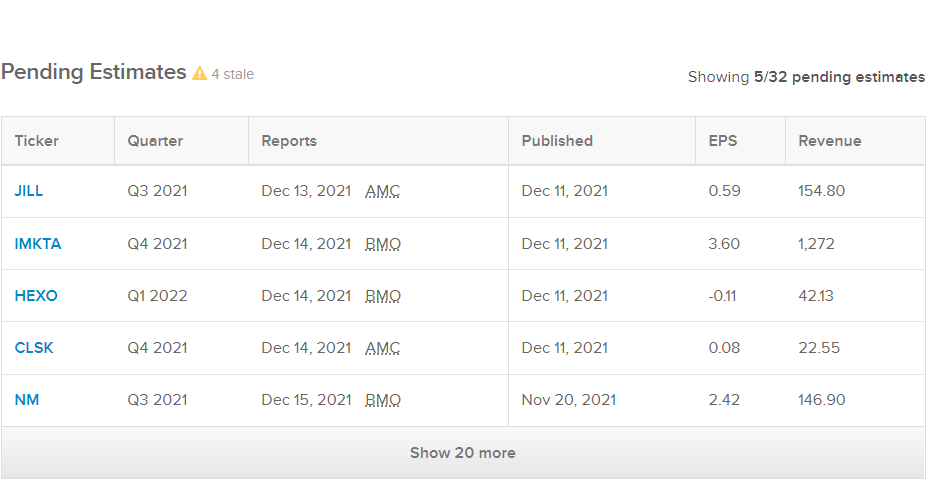
ix. stocks: 1889

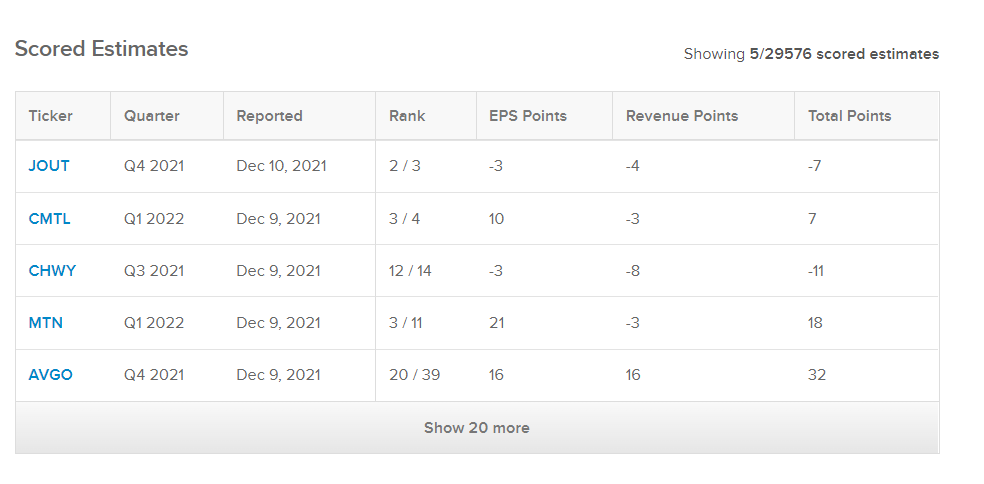
x. pending: 28

**Scrape All covered stock estimates by the analyst**

****

**All pending stock estimates by the analyst**

****

**All scored stock estimates by the analyst**

Graphical user interface, text, application

Description automatically generated

Graphical user interface

Description automatically generated with medium confidence

Analyst.csv



Logo, company name

Description automatically generated

**Importing data into MySQL**

**1. Create a database in MySQL named: msa8040\_finalproject\_sql**

/\* drop a database if exists \*/

drop database if exists msa8040\_finalproject\_sql;

/\* create a database with name 'msa8040\_finalproject\_sql' \*/

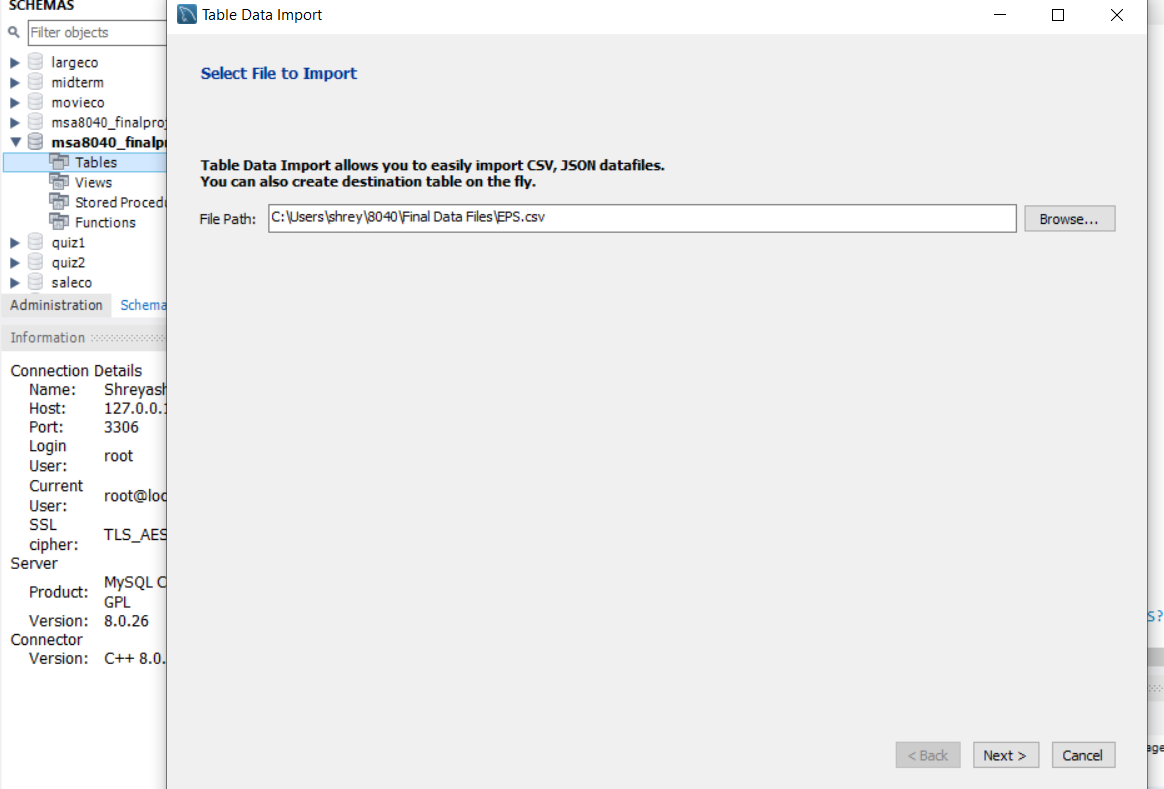
create database msa8040\_finalproject\_sql;

2**. Use the database: msa8040\_finalproject\_sql**

/\* Use database \*/

use msa8040\_finalproject\_sql;

**3. Import all the three csv files; Eps.csv, Company.csv and Analyst.csv in the database: msa8040\_finalproject\_sql using the Table Data Import wizard and follow the next button to the finish button to complete the import.**

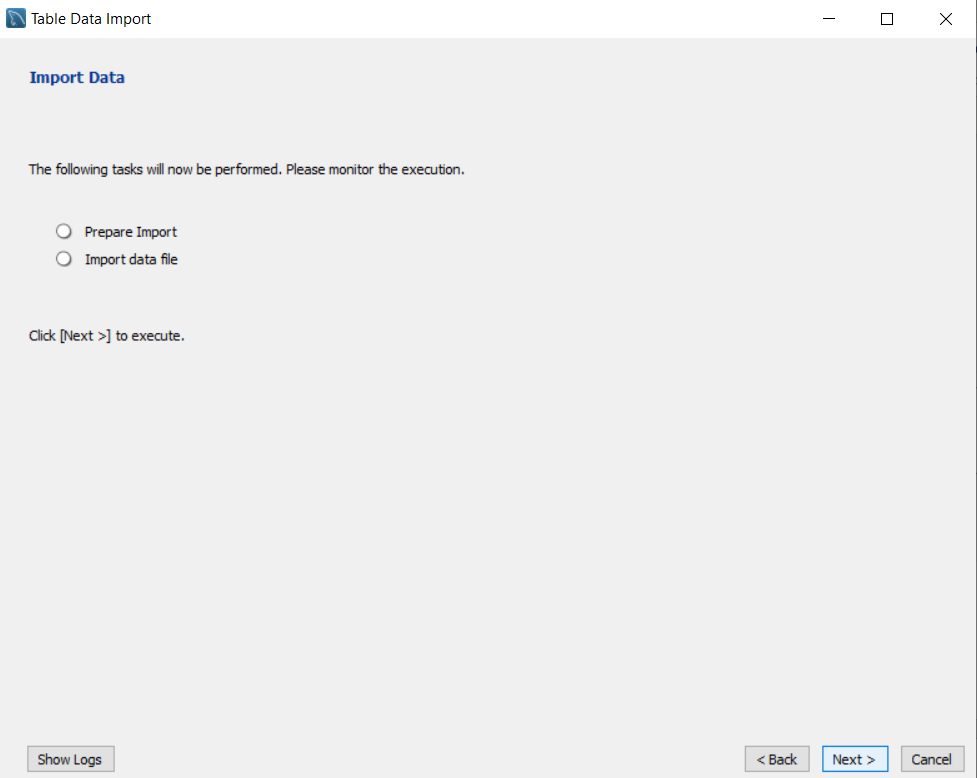


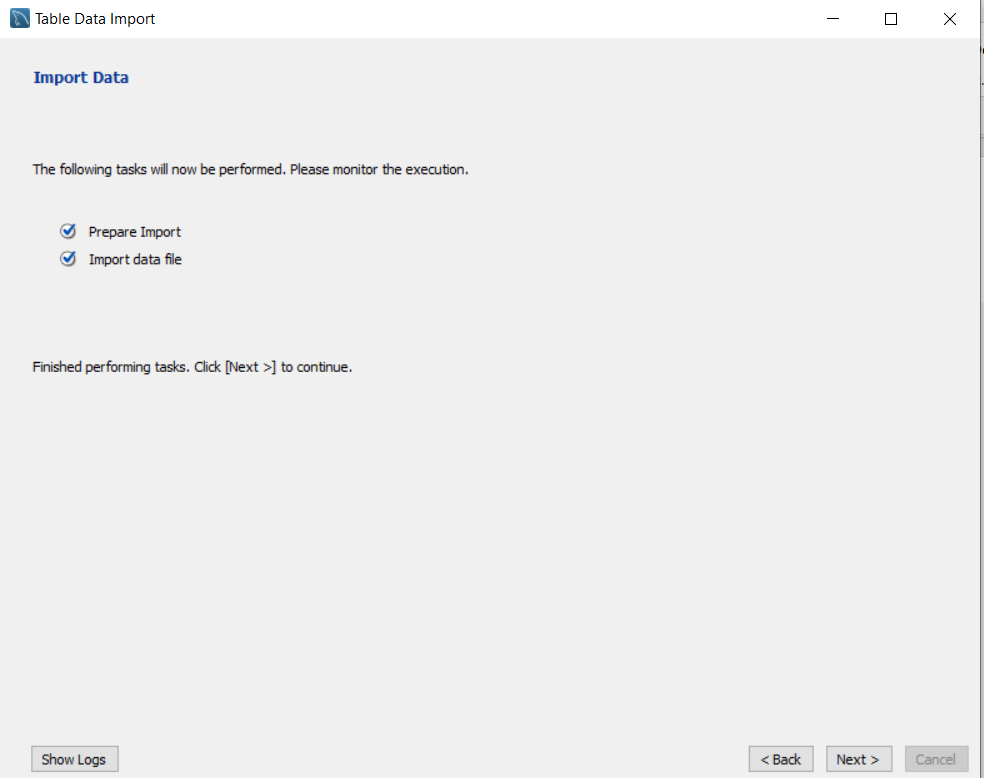
**Graphical user interface, text, application, email

Description automatically generated**

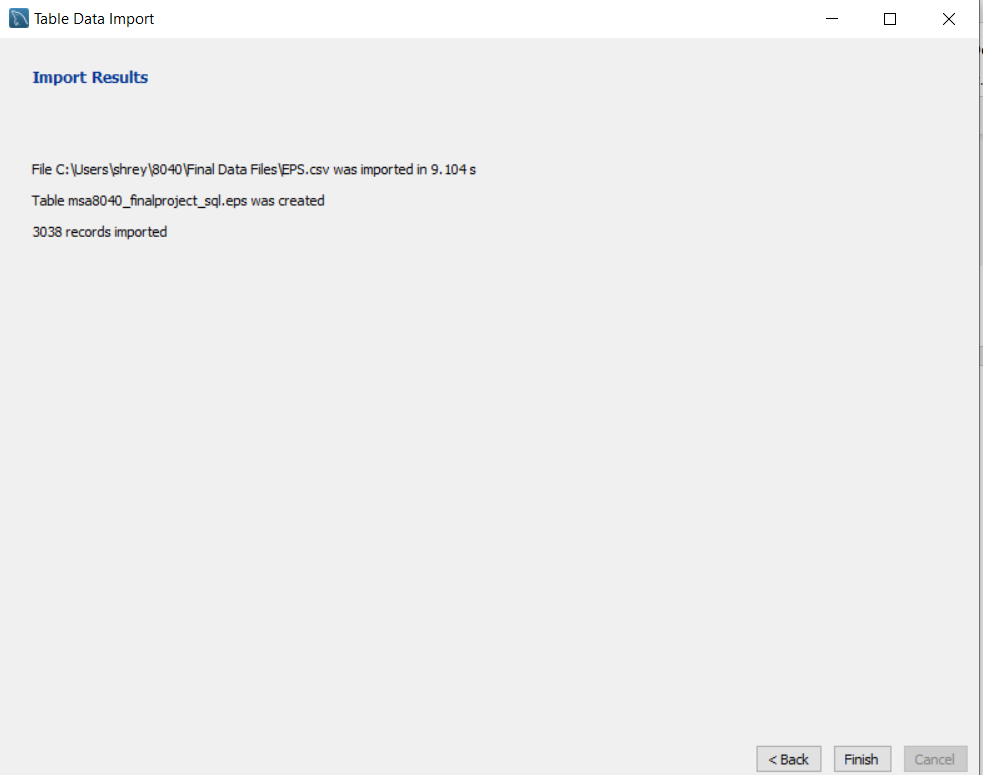
**Table

Description automatically generated**

****

****

**4. Checking for the successful import of the csv files in the database**



**4. Check for the successful import of all the three csv files.**

**/\*Show tables in the database msa8040\_finalproject\_sql \*/**

**show tables;**

Graphical user interface, application

Description automatically generated

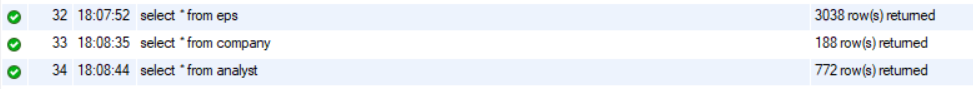
**5. Check for the number of records imported in each table**

/\* Check for number of records in each table \*/

select \* from eps;

select \* from company;

select \* from analyst;

**Results:**

**6. Query the database to retrieve answers to the following questions.**

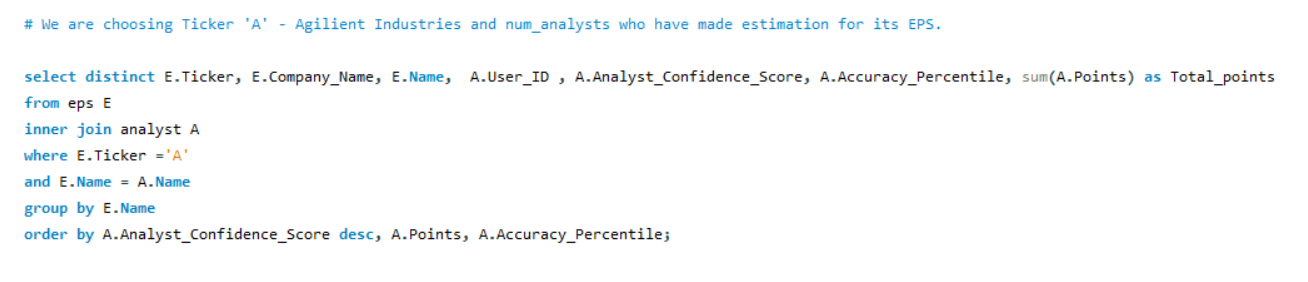
(a) Given a ticker, how many analysts have made estimations for its EPS? Rank them by their confidence score, total points, error rate or accuracy percentile?

**A picture containing timeline

Description automatically generated**

|  |  |
| --- | --- |
| **Adobe Systems Inc.** | **140** |
| **AbbVie Inc.** | **137** |
| **American Airlines Group Inc.** | **136** |
| **Apple Inc.** | **136** |
| **Lululemon Athletica Inc.** | **136** |
| **Autodesk, Inc.** | **129** |
| **Alliance Data Systems Corporation** | **118** |
| **Agilent Technologies Inc.** | **115** |
| **Abbott Laboratories** | **112** |
| **Accenture plc** | **111** |
| **Archer Daniels Midland Company** | **96** |
| **Analog Devices Inc.** | **82** |
| **Axon Enterprise, Inc.** | **78** |
| **Alcoa Corp.** | **75** |
| **Automatic Data Processing, Inc.** | **73** |
| **Applied Optoelectronics, Inc.** | **72** |
| **Atlas Air Worldwide Holdings Inc.** | **72** |
| **Abiomed Inc.** | **71** |
| **ACADIA Pharmaceuticals, Inc.** | **70** |
| **American Eagle Outfitters, Inc.** | **61** |
| **AcelRx Pharmaceuticals, Inc.** | **51** |
| **ADTRAN Inc.** | **50** |
| **Advance Auto Parts Inc.** | **49** |
| **Aaron's, Inc.** | **48** |
| **Advanced Energy Industries, Inc.** | **46** |
| **AAC Holdings, Inc.** | **44** |
| **Acacia Communications, Inc.** | **44** |
| **Axcelis Technologies, Inc.** | **42** |
| **Aurora Cannabis, Inc.** | **41** |
| **Asbury Automotive Group, Inc.** | **38** |
| **AECOM Technology Corporation** | **36** |
| **Addus HomeCare Corporation** | **35** |
| **ACI Worldwide, Inc.** | **35** |
| **American Equity Investment Life Holding Co.** | **35** |
| **Abeona Therapeutics Inc.** | **34** |
| **ACCO Brands Corporation** | **34** |
| **AmerisourceBergen Corporation** | **33** |
| **ABM Industries Inc.** | **32** |
| **Acorda Therapeutics, Inc.** | **31** |
| **Adaptimmune Therapeutics plc** | **31** |
| **Acadia Healthcare Company, Inc.** | **30** |
| **Adamas Pharmaceuticals, Inc.** | **30** |
| **Ameren Corporation** | **27** |
| **Aceto Corporation** | **24** |
| **Aegion Corporation** | **24** |
| **Acacia Research Corporation** | **24** |
| **Advanced Disposal Services, Inc.** | **22** |
| **AAON Inc.** | **21** |
| **Agnico Eagle Mines Ltd.** | **19** |
| **Arcosa, Inc.** | **8** |

**/\* We are choosing Ticker 'A' - Agilient Industries and num\_analysts who have made estimation for its EPS. \*/**

****

select distinct E.Ticker, E.Company\_Name, E.Name, A.User\_ID , A.Analyst\_Confidence\_Score, A.Accuracy\_Percentile, sum(A.Points) as Total\_points

from eps E

inner join analyst A

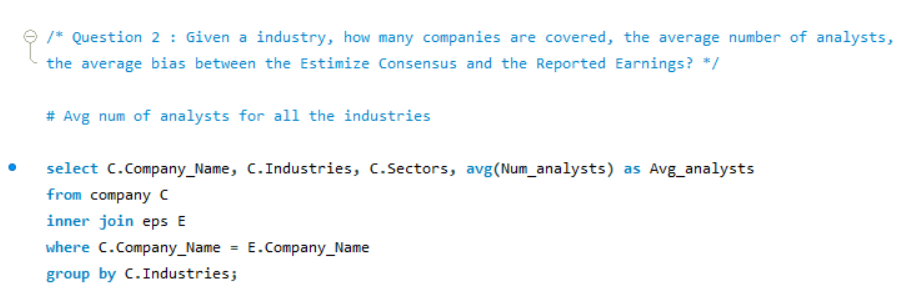
where E.Ticker ='A'

and E.Name = A.Name

group by E.Name

order by A.Analyst\_Confidence\_Score desc, A.Points, A.Accuracy\_Percentile;

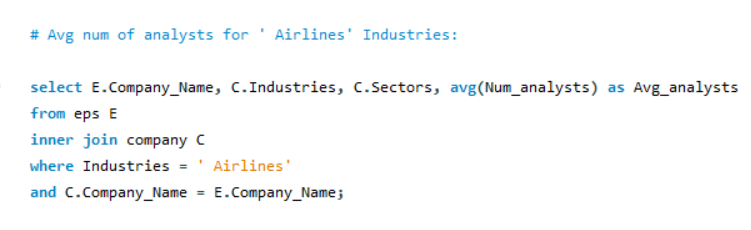
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| A | Agilent Technologies Inc. | arjunav | arjunav | 8.9 | - | 0 |
| A | Agilent Technologies Inc. | Zilvinas Speteliunas | Spekoliunas | 8.8 | - | 0 |
| A | Agilent Technologies Inc. | Richard Wickstrom | RWICK27 | 8.7 | - | 0 |
| A | Agilent Technologies Inc. | Saulius | Sasas | 8.5 | - | 0 |
| A | Agilent Technologies Inc. | serp | serp | 8.5 | - | 0 |
| A | Agilent Technologies Inc. | New\_Moon | New\_Moon | 8.4 | - | 0 |
| A | Agilent Technologies Inc. | SM\_Georg | SM\_Georg | 8.4 | - | 0 |
| A | Agilent Technologies Inc. | Bill | BillB1210 | 8.4 | 59% | 100 |
| A | Agilent Technologies Inc. | dalton | dalton | 8.4 | 67% | 87 |
| A | Agilent Technologies Inc. | Shawn P. Cooney | shawncooney | 8.4 | 57% | 502 |
| A | Agilent Technologies Inc. | panda\_w | panda\_w | 8.3 | 39% | -36 |
| A | Agilent Technologies Inc. | Analyst\_76071 | Analyst\_76071 | 8.3 | 78% | -2 |
| A | Agilent Technologies Inc. | Ben Jen | benjenholdings | 8.3 | 60% | 46 |
| A | Agilent Technologies Inc. | Philip Mascherino | Megaannum | 8.2 | 52% | 448 |
| A | Agilent Technologies Inc. | RmznKrkmz | RmznKrkmz | 8.1 | - | 0 |
| A | Agilent Technologies Inc. | Ram | rama2050 | 8.1 | - | 0 |
| A | Agilent Technologies Inc. | Rob Ford | Old\_Mayor | 8.1 | 56% | 12 |
| A | Agilent Technologies Inc. | Ron | R\_C12R | 8.1 | 55% | 304 |
| A | Agilent Technologies Inc. | Analyst\_783524 | Analyst\_783524 | 8 | - | 0 |
| A | Agilent Technologies Inc. | Analyst\_5795851 | Analyst\_5795851 | 8 | - | 0 |
| A | Agilent Technologies Inc. | kerry burrell | kerryburrell | 8 | 59% | 8 |
| A | Agilent Technologies Inc. | Frederick Tremblay | QuantTrader007 | 7.9 | - | 0 |
| A | Agilent Technologies Inc. | Don Pagach | dpdon985 | 7.9 | 50% | 11 |
| A | Agilent Technologies Inc. | larrylai | larrylai | 7.9 | 59% | 6 |
| A | Agilent Technologies Inc. | Analyst\_1536505 | Analyst\_1536505 | 7.9 | 57% | 302 |
| A | Agilent Technologies Inc. | Roman Novacek | Raven | 7.8 | - | 0 |
| A | Agilent Technologies Inc. | PCLA | PCLA | 7.8 | - | 0 |
| A | Agilent Technologies Inc. | Analyst\_1342063 | Analyst\_1342063 | 7.8 | - | 0 |
| A | Agilent Technologies Inc. | AAFED | AAFED | 7.8 | 47% | 4 |
| A | Agilent Technologies Inc. | Analyst\_2777191 | Analyst\_2777191 | 7.7 | - | 0 |
| A | Agilent Technologies Inc. | maggieb | maggieb | 7.7 | - | 0 |
| A | Agilent Technologies Inc. | Analyst\_8113162 | Analyst\_8113162 | 7.6 | - | 0 |
| A | Agilent Technologies Inc. | Analyst\_6508066 | Analyst\_6508066 | 7.5 | - | 0 |
| A | Agilent Technologies Inc. | Madmoni | Madmoni | 7.4 | 49% | 334 |
| A | Agilent Technologies Inc. | Analyst\_3656041 | Analyst\_3656041 | 7.3 | - | 0 |
| A | Agilent Technologies Inc. | Aitvaras | Aitvaras | 7.3 | - | 0 |
| A | Agilent Technologies Inc. | Analyst\_8202295 | Analyst\_8202295 | 7.3 | 63% | 145 |
| A | Agilent Technologies Inc. | Analyst\_290705 | Analyst\_290705 | 7.2 | - | 0 |
| A | Agilent Technologies Inc. | miltrltr | miltrltr | 7.1 | - | 0 |
| A | Agilent Technologies Inc. | PM | PM | 7.1 | 11% | -2340 |
| A | Agilent Technologies Inc. | Analyst\_2647208 | Analyst\_2647208 | 6.8 | - | 0 |
| A | Agilent Technologies Inc. | Neil\_Martin | Neil\_Martin | 6.8 | 51% | 3 |
| A | Agilent Technologies Inc. | Ketamine\_Fund | Ketamine\_Fund | 6.7 | - | 0 |
| A | Agilent Technologies Inc. | SweUbbe | SweUbbe | 6.7 | 83% | 35 |
| A | Agilent Technologies Inc. | Sentinel | Sentinel | 6.6 | - | 0 |
| A | Agilent Technologies Inc. | Analyst\_9118472 | Analyst\_9118472 | 6.4 | - | 0 |
| A | Agilent Technologies Inc. | Analyst\_6127515 | Analyst\_6127515 | 6.1 | - | 0 |
| A | Agilent Technologies Inc. | JABOM\_AM | JABOM\_AM | 5.8 | - | 0 |
| A | Agilent Technologies Inc. | schmidtke\_jake | schmidtke\_jake | 4.1 | - | 0 |
| A | Agilent Technologies Inc. | Stephen Unger | Needham\_38 | 3.7 | 23% | -284 |

(b) Given a industry, how many companies are covered, the average number of analysts, the average bias between the Estimize Consensus and the Reported Earnings?

Company\_Name Industries Sectors Avg\_analysts

|  |  |  |  |
| --- | --- | --- | --- |
| Agilent Technologies Inc. | Life Sciences Tools & Services | Health Care | 429.0000 |
| Alcoa Corp. | Metals & Mining | Materials | 207.2766 |
| AAC Holdings, Inc. | Health Care Providers & Services | Health Care | 56.3373 |
| American Airlines Group Inc. | Airlines | Industrials | 589.0000 |
| Aaron's, Inc. | Specialty Retail | Consumer Discretionary | 253.5663 |
| Applied Optoelectronics, Inc. | Communications Equipment | Information Technology | 283.4098 |
| AAON Inc. | Building Products | Industrials | 23.0000 |
| Atlas Air Worldwide Holdings Inc. | Air Freight & Logistics | Industrials | 126.0000 |
| Axon Enterprise, Inc. | Aerospace & Defense | Industrials | 281.0000 |
| AbbVie Inc. | Pharmaceuticals | Health Care | 346.4454 |
| Abeona Therapeutics Inc. | Biotechnology | Health Care | 98.9949 |
| ABM Industries Inc. | Commercial Services & Supplies | Industrials | 57.3864 |
| Abiomed Inc. | Health Care Equipment & Supplies | Health Care | 316.3005 |
| Arcosa, Inc. | Construction & Engineering | Industrials | 60.8824 |
| Acacia Communications, Inc. | Semiconductors | Information Technology | 189.9346 |
| ACI Worldwide, Inc. | Software | Information Technology | 386.4207 |
| Accenture plc | IT Services | Information Technology | 270.3510 |
| Acacia Research Corporation | Professional Services | Industrials | 63.0000 |
| Archer Daniels Midland Company | Food Products | Consumer Staples | 212.0000 |
| Ameren Corporation | Multi-Utilities | Utilities | 63.0000 |
| American Equity Investment Life Holding Co. | Insurance | Financials | 38.0000 |

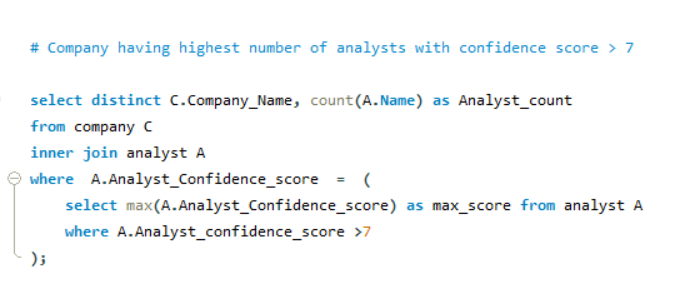
**# Selecting the Industries as “Airlines”**

****

Company\_Name Industries Sectors Avg\_analysts

|  |  |  |  |
| --- | --- | --- | --- |
| American Airlines Group Inc. | Airlines | Industrials | 589.0000 |

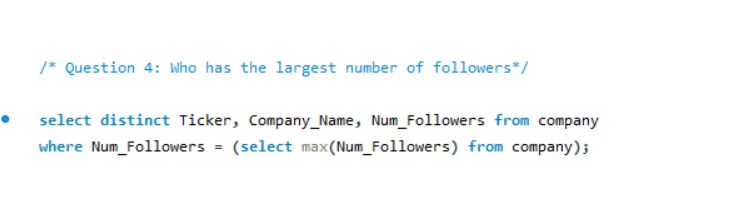
(c) Which company have the largest number of analysts with confidence score greater than 7?



|  |  |
| --- | --- |
| Agilent Technologies Inc. | 188 |

Company\_Name Analyst\_count

|  |  |
| --- | --- |
| Agilent Technologies Inc. | 188 |

(d) Who has the largest number of followers?

Ticker Company\_Name Num\_Followers

|  |  |  |
| --- | --- | --- |
| ABBV | AbbVie Inc. | 650 |