MSA 8040 Data Management for Analytics:

Final Project Report:

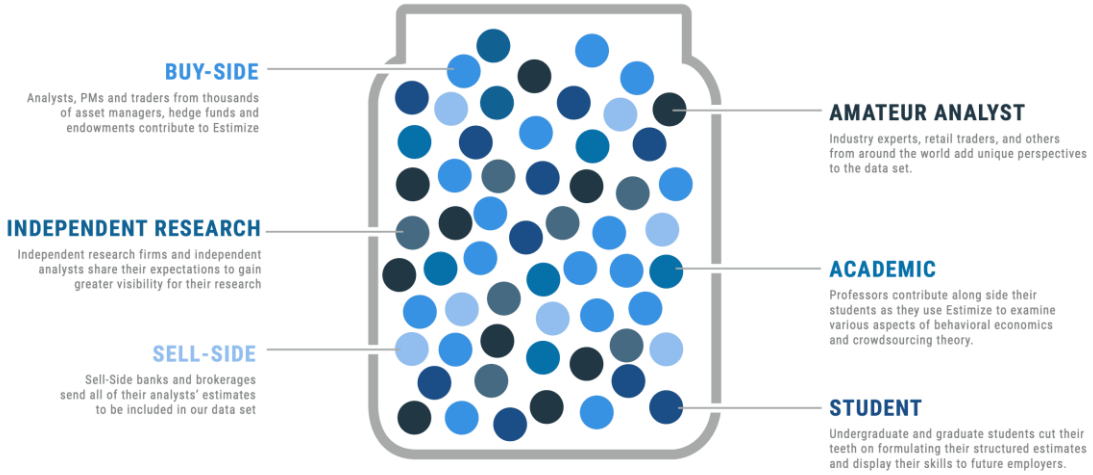
Data Analysis with Estimize.com

Shreyashi Mukhopadhyay (Panther id: 002592906)

Tools used: Selenium, MySQL, Python

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.

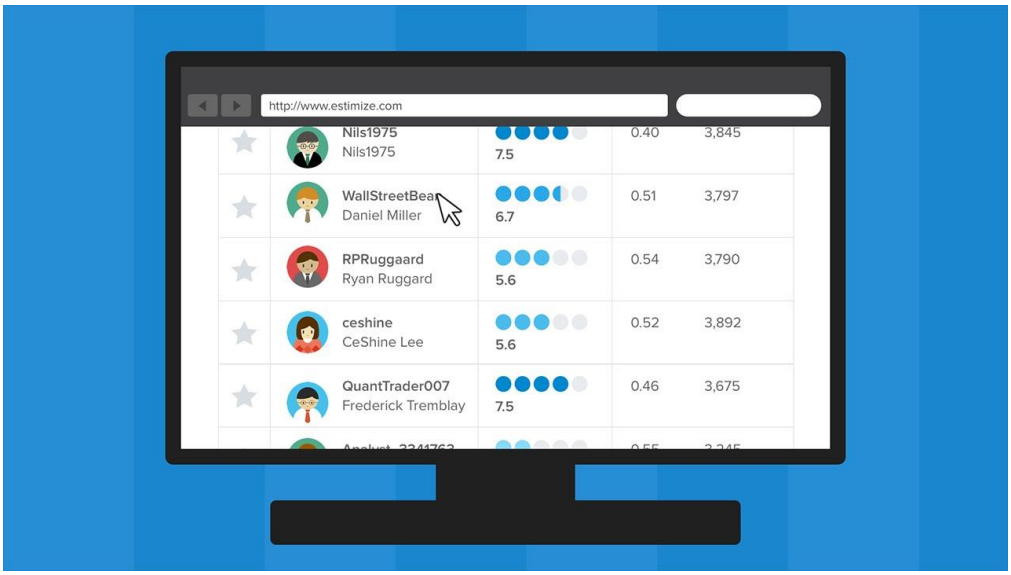


Launched in 2011, Estimize is an open financial estimates platform designed to collect forward looking financial estimates from independent, buy-side, and sell-side analysts, along with those of private investors and academics.

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:

```
import os
import re
import time
import csv
import json
import random
import requests
import pandas as pd
import numpy as np
from datetime import date, timedelta, datetime
import selenium
from selenium import webdriver
from bs4 import BeautifulSoup
```

```
from selenium import webdriver
from selenium.webdriver.common.by import By
from selenium.webdriver.support.ui import WebDriverWait
from selenium.webdriver.support import expected_conditions as EC
from selenium.webdriver.support.ui import Select
from selenium.webdriver.support.select import Select
from selenium.webdriver.chrome.options import Options
# from selenium.webdriver.common.by import By
```

```
pip install selenium==3.141.0
```

```
Requirement already satisfied: selenium==3.141.0 in c:\users\shrey\anaconda3\lib\site-packages (3.141.0)
Requirement already satisfied: urllib3 in c:\users\shrey\anaconda3\lib\site-packages (from selenium==3.141.0) (1.25.11)
Note: you may need to restart the kernel to use updated packages.
```

```
pip install webdriver-manager
```

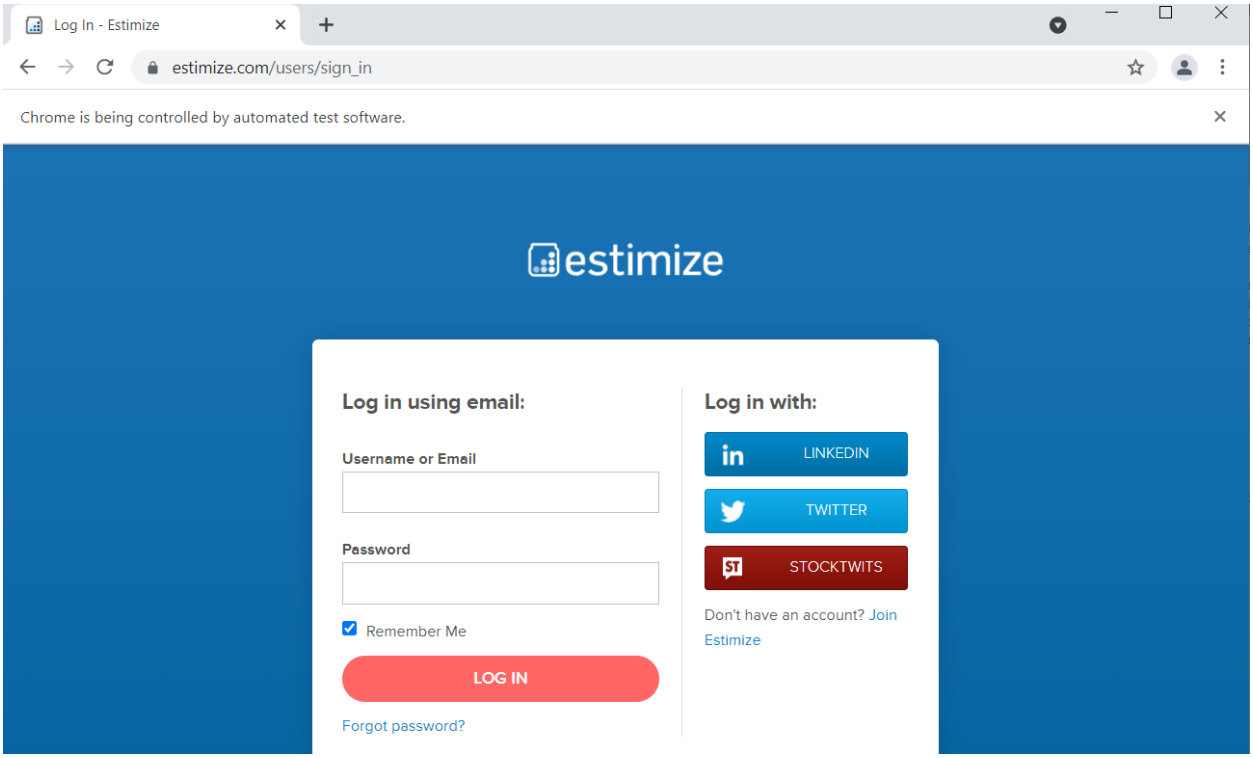
2. Setting up the chrome driver and defining the browser object

Set up the chrome driver by passing some parameters

```
chrome_options = webdriver.ChromeOptions()
chrome_options.headless = False
browser = webdriver.Chrome(executable_path = 'C:/Users/shrey/Downloads/chromedriver_win32/chromedriver.exe',
                           options = chrome_options)
url_head = 'https://www.estimize.com/users/sign_in'
browser.get(url_head)
time.sleep(random.uniform(1,5))
```

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

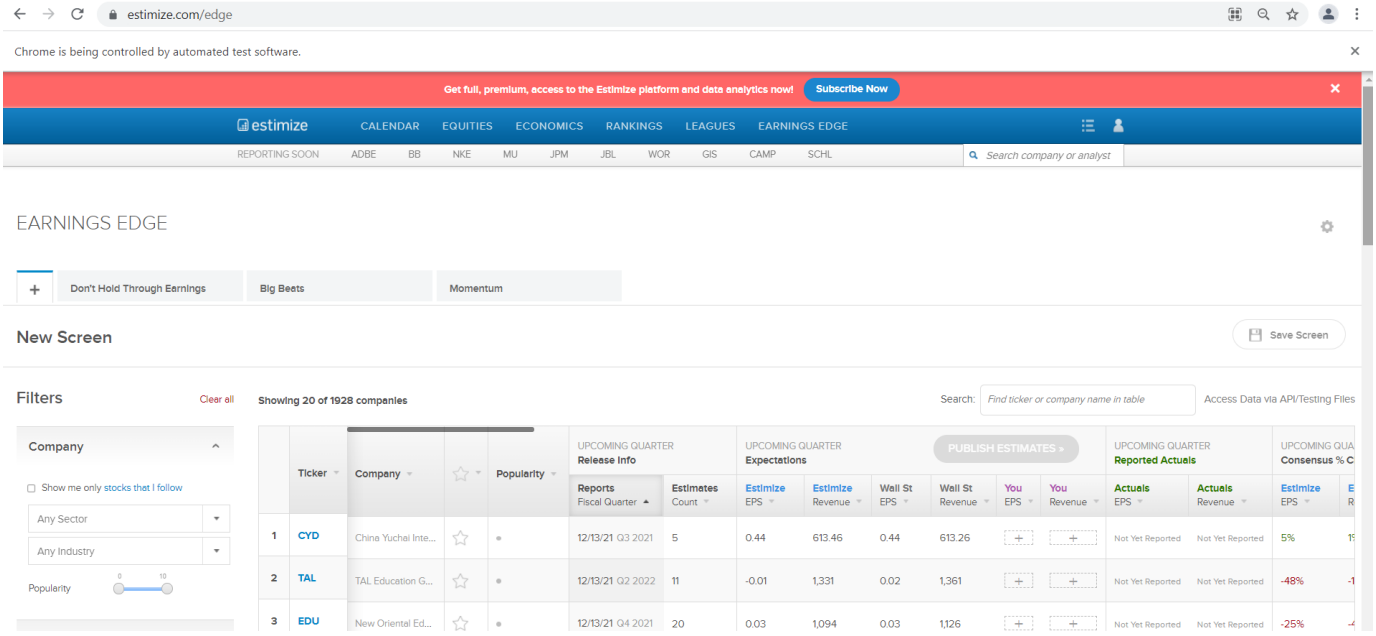
User Authentication

Filling in sign_in page using the username and password cells

```
def sign_in():
    url = 'https://www.estimize.com/users/sign_in'
    browser.get(url)



    email_field = browser.find_element_by_id('user_login')
    password_field = browser.find_element_by_id('user_password')
    submit_btn = browser.find_element_by_xpath('//input[@type="submit"]')



    username = 'smukhopadhyay1@student.gsu.edu'
    password = 'Shreyashi_03'



    email_field.send_keys(username)
    password_field.send_keys(password)
    submit_btn.click()

    time.sleep(1)
```

```
sign_in()
```

4. User page is displayed upon successful login.



5. Ticker and quarter selection.

Ticker and quarter selection ¶

```
# url = 'https://www.glassdoor.com/Overview/Working-at-ADC-Telecommunications-EI_IE1075.11,33.htm'

tickers = ['A', 'AA', 'AACH', 'AAL', 'AAN', 'AAOI', 'AAON', 'AAP', 'AAPL', 'AAWW', 'AAXN','ABBV', 'ABC', 'ABEO', 'ABG','ABM','ABMD', 'ABT',
           'ACA', 'ACAD', 'ACB', 'ACCO', 'LULU', 'ACET', 'ACHC', 'ACIA', 'ACIW', 'ACLS','ACM','ACN','ACOR','ACRX', 'ACTG', 'ADAP',
           'ADBE', 'ADI','ADM','ADMS', 'ADP', 'ADS', 'ADSK', 'ADSW', 'ADTN',  'ADUS', 'AEE', 'AEGN', 'AEIS', 'AEL', 'AEM', 'AEO']

quarters = ['fq1-2020', 'fq2-2020', 'fq3-2020', 'fq4-2020']
```

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:

Extract data for Company table

```
def scrape_overview(browser):

    overviewAll = {}

    try:
        for i in range(len(tickers)):
            for j in range(len(quarters)):
#             url = 'https://www.estimize.com/'+tickers[i]+'/fq3-2021?metric_name=eps&chart=historical'
              url = 'https://www.estimize.com/'+tickers[i]+ '/' + quarters[j]+ '?metric_name=eps&chart=historical'
    #     https://www.estimize.com/a/fq3-2021?metric_name=eps&chart=historical
              browser.get(url)
              time.sleep(1)
              overview = {}
              overview['quarter'] = quarters[j]
              overview['Ticker'] = browser.find_element_by_class_name('release-header-information-title').text
              overview['Company Name'] = browser.find_element_by_class_name('release-header-information-description').text
              sectors = browser.find_element_by_class_name('release-header-information-breadcrumb')
              overview['Sectors'] = sectors.text.split('›')[0]
              overview['Industries'] = sectors.text.split('›')[1]
              numbers = browser.find_elements_by_class_name('summary-sub-title')
              overview['Number of Followers'] = numbers[0].text
              overview['Number of Analysts'] = numbers[1].text

              overviewAll[len(overviewAll)]=overview
    except Exception:
        pass

    return overviewAll
```

Create Company CSV/JSON File

```
company_info = scrape_overview(browser)

with open('Company.json', 'w') as outfile:
    json.dump(company_info, outfile, indent=4)

df=pd.DataFrame(company_info)
df_trans = df.T
df_trans.to_csv('Company.csv')
```

Company.csv

quarter	Ticker	Company_Name	Sectors	Industries	Num_Followers	Num_Analysts
fq1-2020	A	Agilent Technologies Inc.	Health Care	Life Sciences Tools & Services	249	429
fq2-2020	A	Agilent Technologies Inc.	Health Care	Life Sciences Tools & Services	249	429
fq3-2020	A	Agilent Technologies Inc.	Health Care	Life Sciences Tools & Services	249	429
fq4-2020	A	Agilent Technologies Inc.	Health Care	Life Sciences Tools & Services	249	429
fq1-2020	AA	Alcoa Corp.	Materials	Metals & Mining	167	257
fq2-2020	AA	Alcoa Corp.	Materials	Metals & Mining	167	257
fq3-2020	AA	Alcoa Corp.	Materials	Metals & Mining	167	257
fq4-2020	AA	Alcoa Corp.	Materials	Metals & Mining	167	257
fq1-2020	AACH	AAC Holdings, Inc.	Health Care	Health Care Providers & Services	13	33
fq2-2020	AACH	AAC Holdings, Inc.	Health Care	Health Care Providers & Services	13	33
fq3-2020	AACH	AAC Holdings, Inc.	Health Care	Health Care Providers & Services	13	33
fq4-2020	AACH	AAC Holdings, Inc.	Health Care	Health Care Providers & Services	13	33
fq1-2020	AAL	American Airlines Group Inc.	Industrials	Airlines	618	589
fq2-2020	AAL	American Airlines Group Inc.	Industrials	Airlines	618	589
fq3-2020	AAL	American Airlines Group Inc.	Industrials	Airlines	618	589
fq4-2020	AAL	American Airlines Group Inc.	Industrials	Airlines	618	589
fq1-2020	AAN	Aaron's, Inc.	Consumer Discretionary	Specialty Retail	47	114
fq2-2020	AAN	Aaron's, Inc.	Consumer Discretionary	Specialty Retail	47	114
fq3-2020	AAN	Aaron's, Inc.	Consumer Discretionary	Specialty Retail	47	114
fq4-2020	AAN	Aaron's, Inc.	Consumer Discretionary	Specialty Retail	47	114
fq1-2020	AAOI	Applied Optoelectronics, Inc.	Information Technology	Communications Equipment	196	333
fq2-2020	AAOI	Applied Optoelectronics, Inc.	Information Technology	Communications Equipment	196	333
fq3-2020	AAOI	Applied Optoelectronics, Inc.	Information Technology	Communications Equipment	196	333
fq4-2020	AAOI	Applied Optoelectronics, Inc.	Information Technology	Communications Equipment	196	333

Building the EPS table:

This table contains the EPS information, including:

i. Reported Earnings

ii. Estimize Consensus

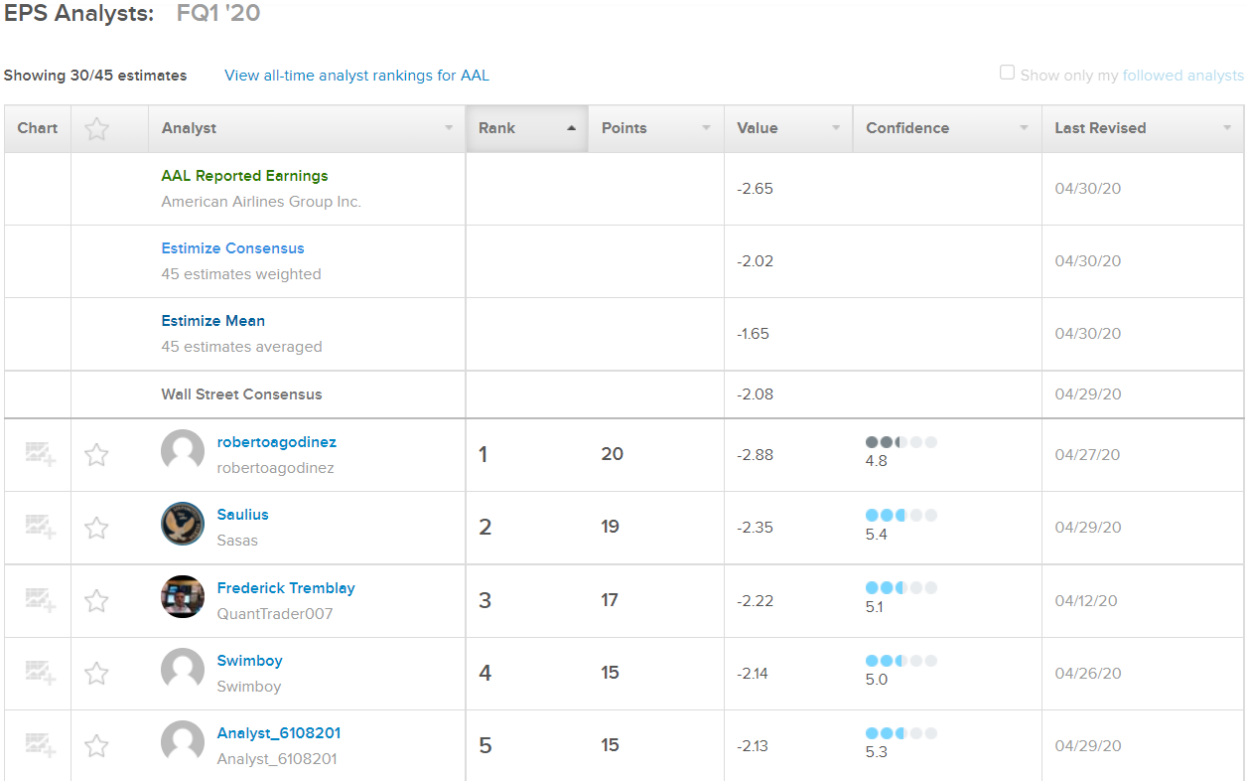
iii. Estimize Mean

iv. Wall Street Consensus

v. 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



EPS Analysts: FQ1 '20

Showing 30/45 estimates View all-time analyst rankings for AAL Show only my followed analysts

Chart	☆	Analyst	Rank	Points	Value	Confidence	Last Revised
		AAL Reported Earnings American Airlines Group Inc.			-2.65		04/30/20
		Estimize Consensus 45 estimates weighted			-2.02		04/30/20
		Estimize Mean 45 estimates averaged			-1.65		04/30/20
		Wall Street Consensus			-2.08		04/29/20
	☆	robertoagodinez robertoagodinez	1	20	-2.88	4.8	04/27/20
	☆	Saulius Sasas	2	19	-2.35	5.4	04/29/20
	☆	Frederick Tremblay QuantTrader007	3	17	-2.22	5.1	04/12/20
	☆	Swimboy Swimboy	4	15	-2.14	5.0	04/26/20
	☆	Analyst_6108201 Analyst_6108201	5	15	-2.13	5.3	04/29/20

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

Extract data for EPS table

```
def scrape_eps(browser):
    eps = {}
#     tickers1 = ['A', 'AA' ]
#     try:
    for i in range(len(tickers)):
        for j in range(len(quarters)):
            url = 'https://www.estimize.com/'+tickers[i]+ '/' + quarters[j] + '?metric_name=eps&chart=historical'
            browser.get(url)
            time.sleep(1)
            analysts_generic = browser.find_elements_by_class_name('estimates-tbl-consensus-column')
            values_generic = browser.find_elements_by_class_name('estimates-tbl-consensus-eps')

            for generic_count in range(len(analysts_generic)):
                eps_generic = {}
                eps_generic['quarter'] = quarters[j]
                eps_generic['Ticker'] = browser.find_element_by_class_name('release-header-information-title').text
                eps_generic['Company Name'] = browser.find_element_by_class_name('release-header-information-description').text
                eps_generic['Name'] = analysts_generic[generic_count].text.split('\n')[0]
                eps_generic['Type'] = "Generic"
                eps_generic['Estimated Value'] = values_generic[generic_count].text
                eps [len(eps)] = eps_generic

            analysts = browser.find_elements_by_class_name('username')
            analysts_values = browser.find_elements_by_class_name('estimates-tbl-eps')

            for analyst_count in range(len(analysts)):
                eps_Analyst = {}
                eps_Analyst['quarter'] = quarters[j]
                eps_Analyst['Ticker'] = browser.find_element_by_class_name('release-header-information-title').text
                eps_Analyst['Company Name'] = browser.find_element_by_class_name('release-header-information-description').text
                eps_Analyst['Name'] = analysts[analyst_count].text
                eps_Analyst['Type'] = "Analyst"
                eps_Analyst['Estimated Value'] = analysts_values[analyst_count].text

                eps [len(eps)] = eps_Analyst

    return eps
```

Create EPS information Data/JSON File

```
eps_info = scrape_eps(browser)

with open('EPS.json', 'w') as outfile:
    json.dump(eps_info, outfile, indent=4)

df=pd.DataFrame(eps_info)
df_trans = df.T
df_trans.to_csv('EPS.csv')
```

Eps.csv

A	B	C	D	E	F
quarter	Ticker	Company_Name	Name	Type	Estimated_Value
fq1-2020	A	Agilent Technologies Inc.	A Reported Earnings	Generic	0.81
fq1-2020	A	Agilent Technologies Inc.	A Guidance	Generic	0.81
fq1-2020	A	Agilent Technologies Inc.	Estimize Consensus	Generic	0.83
fq1-2020	A	Agilent Technologies Inc.	Estimize Mean	Generic	0.83
fq1-2020	A	Agilent Technologies Inc.	Wall Street Consensus	Generic	0.81
fq1-2020	A	Agilent Technologies Inc.	Stephen Unger	Analyst	0.81
fq1-2020	A	Agilent Technologies Inc.	SweUbbe	Analyst	0.81
fq1-2020	A	Agilent Technologies Inc.	Sentinel	Analyst	0.82
fq1-2020	A	Agilent Technologies Inc.	Analyst_5795851	Analyst	0.82
fq1-2020	A	Agilent Technologies Inc.	Analyst_2647208	Analyst	0.8
fq1-2020	A	Agilent Technologies Inc.	larrylai	Analyst	0.83
fq1-2020	A	Agilent Technologies Inc.	maggieb	Analyst	0.83
fq1-2020	A	Agilent Technologies Inc.	AAFED	Analyst	0.83
fq1-2020	A	Agilent Technologies Inc.	Analyst_6098124	Analyst	0.83
fq1-2020	A	Agilent Technologies Inc.	miltrltr	Analyst	0.83
fq1-2020	A	Agilent Technologies Inc.	Bill	Analyst	0.83
fq1-2020	A	Agilent Technologies Inc.	Zilvinas Speteliunas	Analyst	0.83
fq1-2020	A	Agilent Technologies Inc.	Analyst_290705	Analyst	0.83
fq1-2020	A	Agilent Technologies Inc.	Rob Ford	Analyst	0.83

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

Stocks Covered

Showing **10 / 2002 stocks covered**

Ticker	Reports	Quarters	Points	Pts/Est	Error Rate	Accuracy
AAPL	Feb 1	30	484	16.1	4.4%	63%
FB	Feb 2	28	546	19.5	6.3%	62%
DIS	Feb 8	27	352	13	12.0%	65%
CRM	Mar 1	26	632	24.3	6.8%	64%
CSCO	Feb 16	26	579	22.3	1.4%	62%
MSFT	Jan 27	26	560	21.5	4.9%	65%
ORCL	Mar 17	25	426	17	2.4%	63%
NKE	Dec 20	25	505	20.2	7.6%	69%
LULU	Mar 30	25	480	19.2	6.1%	68%
CMG	Feb 8	25	384	15.4	7.4%	62%

Show 20 more

All pending stock estimates by the analyst

Pending Estimates ⚠ 4 stale

Showing **5/32 pending estimates**

Ticker	Quarter	Reports	Published	EPS	Revenue
JILL	Q3 2021	Dec 13, 2021 AMC	Dec 11, 2021	0.59	154.80
IMKTA	Q4 2021	Dec 14, 2021 BMO	Dec 11, 2021	3.60	1,272
HEXO	Q1 2022	Dec 14, 2021 BMO	Dec 11, 2021	-0.11	42.13
CLSK	Q4 2021	Dec 14, 2021 AMC	Dec 11, 2021	0.08	22.55
NM	Q3 2021	Dec 15, 2021 BMO	Nov 20, 2021	2.42	146.90

Show 20 more

All scored stock estimates by the analyst

Scored Estimates

Showing **5/29576 scored estimates**

Ticker	Quarter	Reported	Rank	EPS Points	Revenue Points	Total Points
JOUT	Q4 2021	Dec 10, 2021	2 / 3	-3	-4	-7
CMTL	Q1 2022	Dec 9, 2021	3 / 4	10	-3	7
CHWY	Q3 2021	Dec 9, 2021	12 / 14	-3	-8	-11
MTN	Q1 2022	Dec 9, 2021	3 / 11	21	-3	18
AVGO	Q4 2021	Dec 9, 2021	20 / 39	16	16	32

Show 20 more

Extract data for Analyst table

```
def scrape_analysts_info(browser):

    analysts_url = scrape_analysts_list(browser)
    analyst_Profile = ['Error rate', 'Accuracy Percentile', 'Points', 'Points/Estimate', 'Stocks', 'Pending']
    analysts_Overview_final = {}
    for i in range(len(analysts_url)):
        url = analysts_url[i]
        browser.get(url)
        time.sleep(1)
#         print(url)
        analyst_Overview = {}
        try:
            analyst_Overview['Name'] = browser.find_element_by_class_name('profile-display-name').text
            analyst_Overview['User ID'] = browser.find_element_by_class_name('profile-username').text
            # bio = browser.find_elements_by_class_name('profile-bio-categorizations')
            analyst_Overview['roles'] = browser.find_element_by_class_name('profile-bio-categorizations').text
            date = browser.find_element_by_class_name('profile-activity-stats').text.split (' ')
            analyst_Overview['Join Date'] = date[2] + ' ' + date[3]
            analyst_Overview['Analyst Confidence Score'] = browser.find_element_by_class_name('value').text

            profile_stat = browser.find_elements_by_class_name('profile-stat')
            for i in range (len(profile_stat)):
                analyst_Overview[analyst_Profile[i]] = profile_stat[i].text

            analysts_Overview_final[len(analysts_Overview_final)] = analyst_Overview
        except Exception:
            pass

    return analysts_Overview_final
```

Create Analyst Data/JSON File

```
analysts_info = scrape_analysts_info(browser)

with open('Analyst.json', 'w') as outfile:
    json.dump(analysts_info, outfile, indent=4)

df=pd.DataFrame(analysts_info)
df_trans = df.T
df_trans.to_csv('Analyst.csv')
```

```
799
```

Analyst.csv

Name	User_ID	roles	Join_Date	Analyst_Confidence_Score	Error_rate	Accuracy_Percentile	Points	Points_Estimate	Stocks	Pending
Rob Ford	Old_Mayor	Financial Professional Buy Side Asset Manager	15-Jun	8.1	15.30%	56%	3,728	12	310	7
SweUbbe	SweUbbe	Non Professional Financials Professional Services	17-Oct	6.7	22.90%	83%	35	35	1	0
Stephen Unger	Needham_38	Financial Professional Sell Side Broker	19-Jan	3.7	31.00%	23%	-71	-7.9	9	7
Sentinel	Sentinel	Financial Professional Independent Independent Research	18-Aug	6.6	-	-	-	-	-	0
Analyst_5795851	Analyst_5795851		19-Apr	8	-	-	-	-	-	0
Analyst_2647208	Analyst_2647208	Non Professional Financials Capital Markets	16-Oct	6.8	-	-	-	-	-	0
larrylai	larrylai	Financial Professional Independent Independent Research	14-Mar	7.9	15.50%	59%	2,636	11.8	223	12
maggieb	maggieb	Non Professional Consumer Discretionary Media	18-May	7.7	-	-	-	-	-	0
AAFED	AAFED	Non Professional Information Technology IT Services	19-Feb	7.8	17.80%	47%	1,706	5.6	302	201
Analyst_6098124	Analyst_6098124	Non Professional Other Other	20-Feb	N/A	-	-	-	-	-	0
miltrltr	miltrltr	Non Professional Financials Real Estate Development	14-Mar	7.1	-	-	-	-	-	0
Bill	BillB1210	Non Professional Financials Professional Services	14-Jul	8.4	16.50%	59%	25,628	13.6	1,880	24
Zilvinas Speteliunas	Spekoliunas	Financials	16-Oct	8.8	-	-	-	-	-	0
Analyst_290705	Analyst_290705	Non Professional Other Other	20-Feb	7.2	-	-	-	-	-	0
Frederick Tremblay	QuantTrader007	Financial Professional Buy Side Asset Manager	14-Apr	7.9	-	-	-	-	-	0
kerry burrell	kerryburrell	Non Professional Financials Real Estate Development	13-Jun	8	10.30%	59%	2,730	16.4	166	5
New_Moon	New_Moon	Non Professional Academia	20-Jan	8.4	-	-	-	-	-	0
PM	PM	Financial Professional Buy Side Mutual Fund	16-Jul	7.1	22.90%	11%	-585	-17.2	34	0
Philip Mascherino	Megaannum	Financial Professional Independent Independent Research	14-Feb	8.2	10.50%	52%	448	10.9	41	4
serp	serp	Financial Professional Independent Independent Research	17-Mar	8.5	-	-	-	-	-	0
Ram	rama2050	Non Professional Information Technology IT Services	16-Apr	8.1	-	-	-	-	-	0
Richard Wickstrom	RWICK27	Non Professional Information Technology IT Services	15-Jul	8.7	-	-	-	-	-	0
SM_Georg	SM_Georg	Non Professional Student	14-Jan	8.4	-	-	-	-	-	0
Saulius	Sasas	Financial Professional Independent Independent Research	17-May	8.5	-	-	-	-	-	1
Analyst_7792487	Analyst_7792487	Non Professional Other Other	20-Feb	N/A	-	-	-	-	-	0
JABOM_AM	JABOM_AM	Non Professional Student	17-Jul	5.8	-	-	-	-	-	1

Importing data into MySQL

1. Create a database in MySQL named: `msa8040_finalproject_sql`

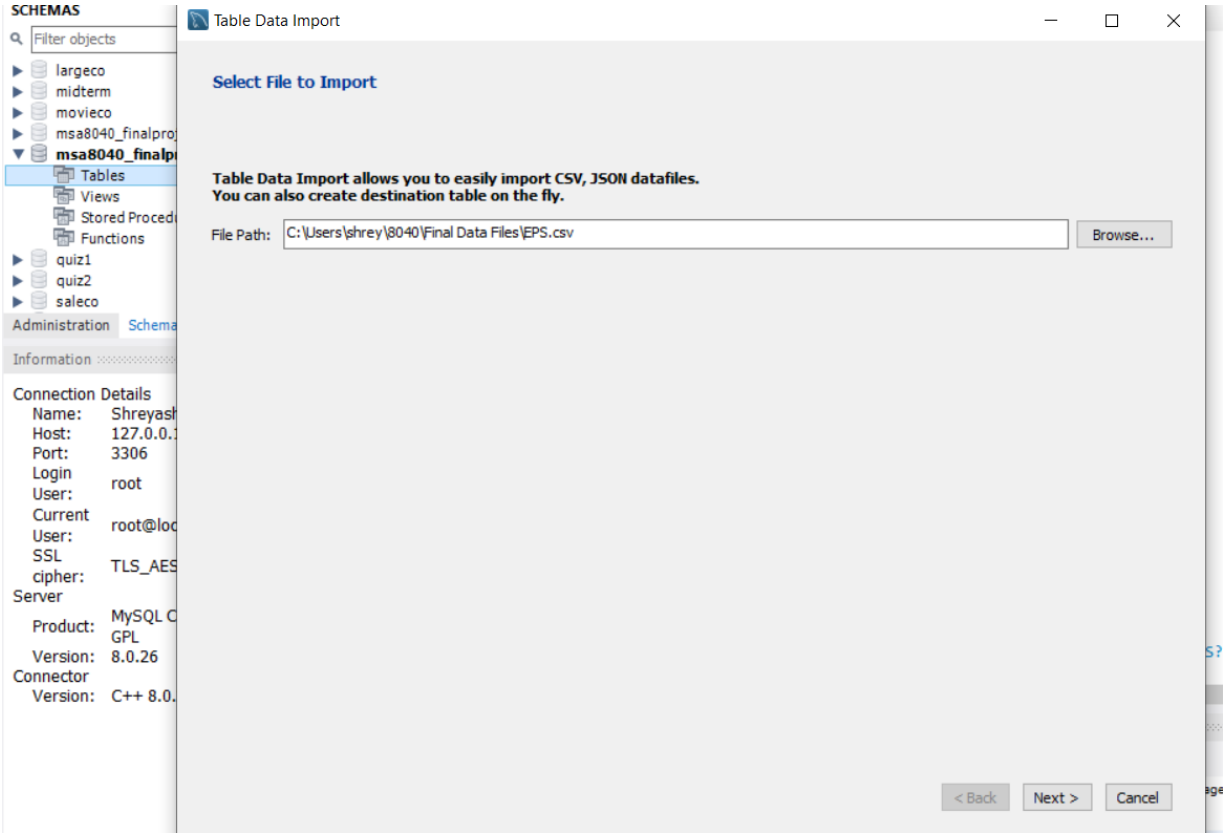
```
/* drop a database if exists */
drop database if exists msa8040_finalproject_sql;
```

```
/* 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.



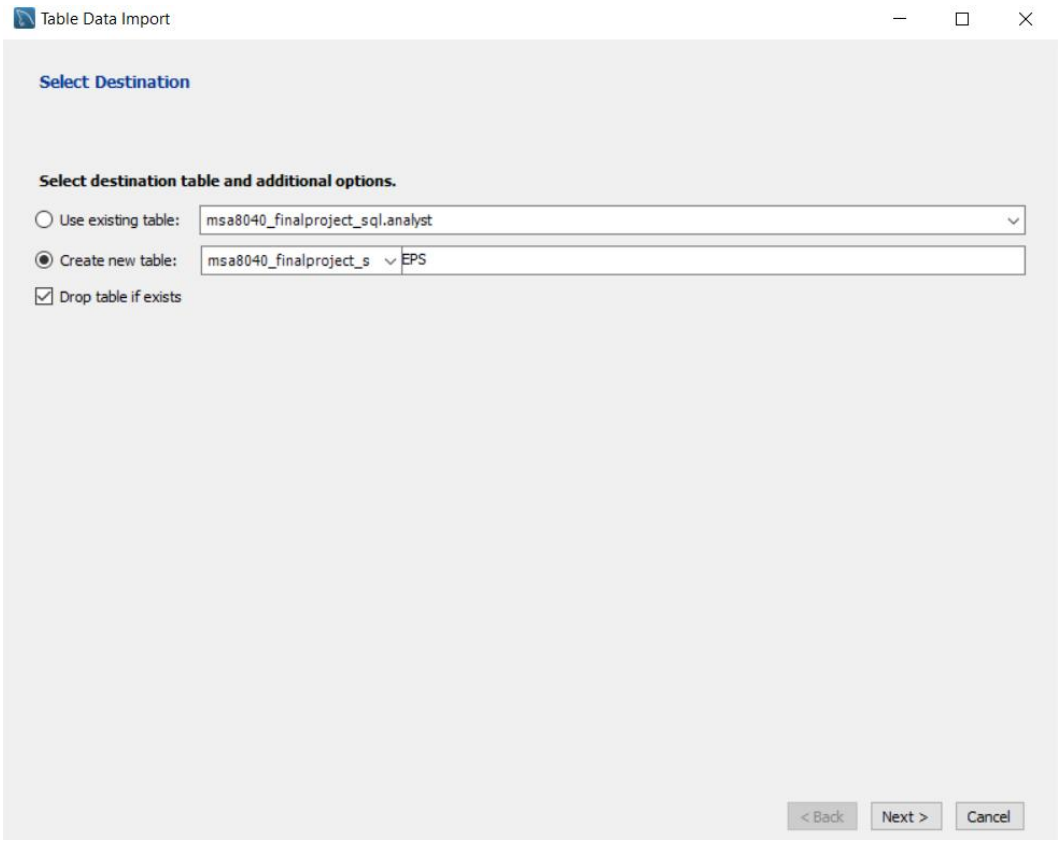


Table Data Import
Select Destination
Select destination table and additional options.
Use existing table:
msa8040_finalproject_sql.analyst
Create new table:
msa8040_finalproject_s
EPS
Drop table if exists
< Back
Next >
Cancel

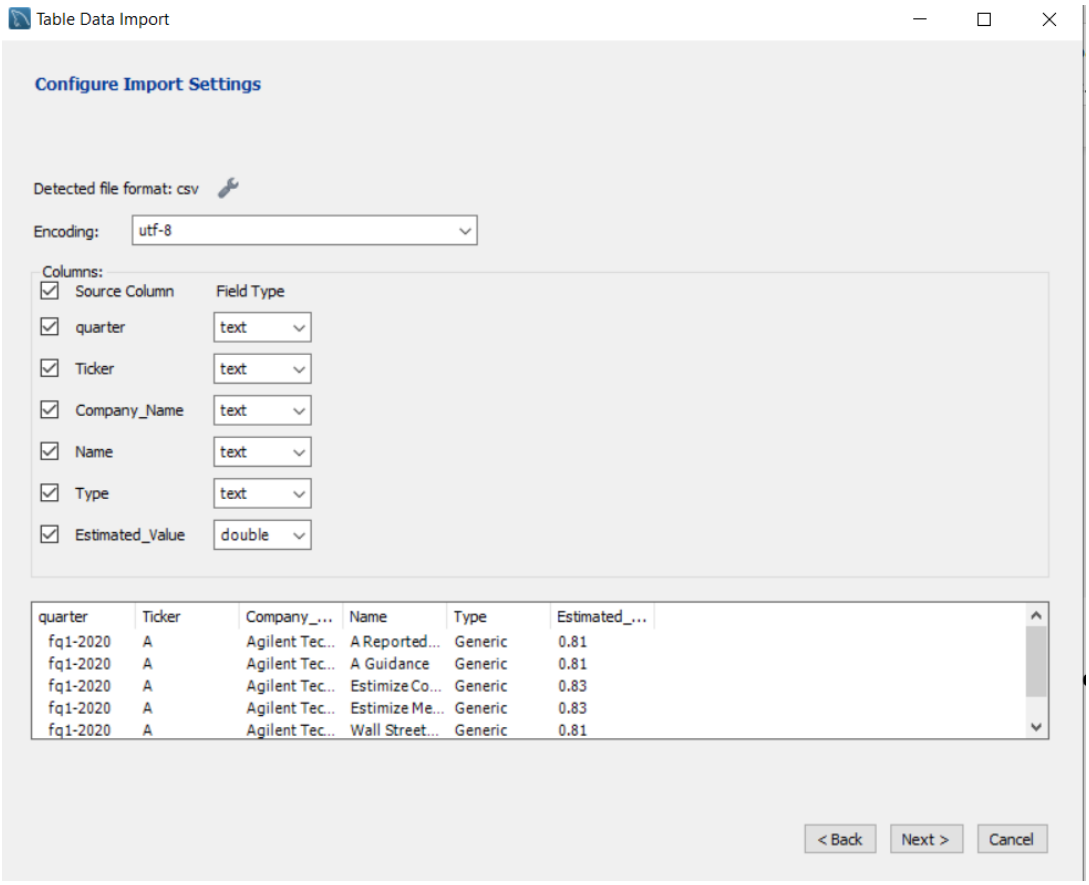


Table Data Import
Configure Import Settings
Detected file format: csv
Encoding:
utf-8
Columns:
Source Column
Field Type
quarter
text
Ticker
text
Company_Name
text
Name
text
Type
text
Estimated_Value
double
quarter
Ticker
Company_...
Name
Type
Estimated_...
fq1-2020
A
Agilent Tec...
A Reported...
Generic
0.81
fq1-2020
A
Agilent Tec...
A Guidance
Generic
0.81
fq1-2020
A
Agilent Tec...
Estimize Co...
Generic
0.83
fq1-2020
A
Agilent Tec...
Estimize Me...
Generic
0.83
fq1-2020
A
Agilent Tec...
Wall Street...
Generic
0.81
< Back
Next >
Cancel

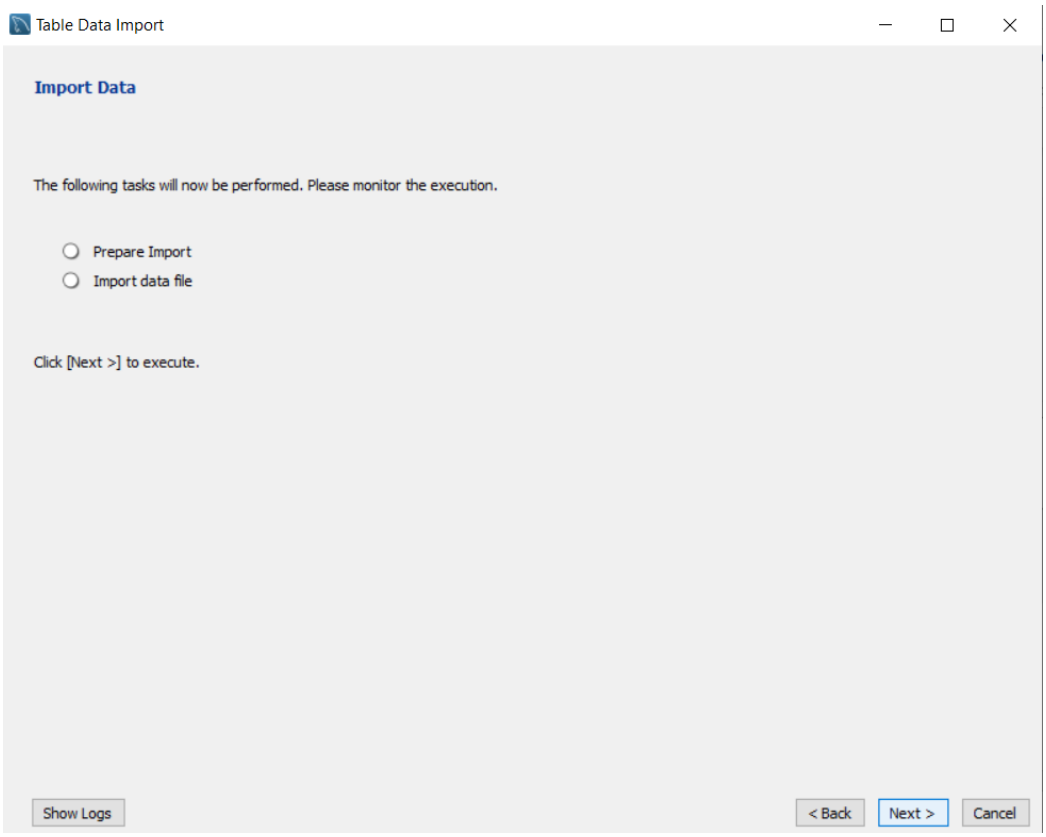


Table Data Import
Import Data
The following tasks will now be performed. Please monitor the execution.
Prepare Import
Import data file
Click [Next >] to execute.
Show Logs
< Back
Next >
Cancel

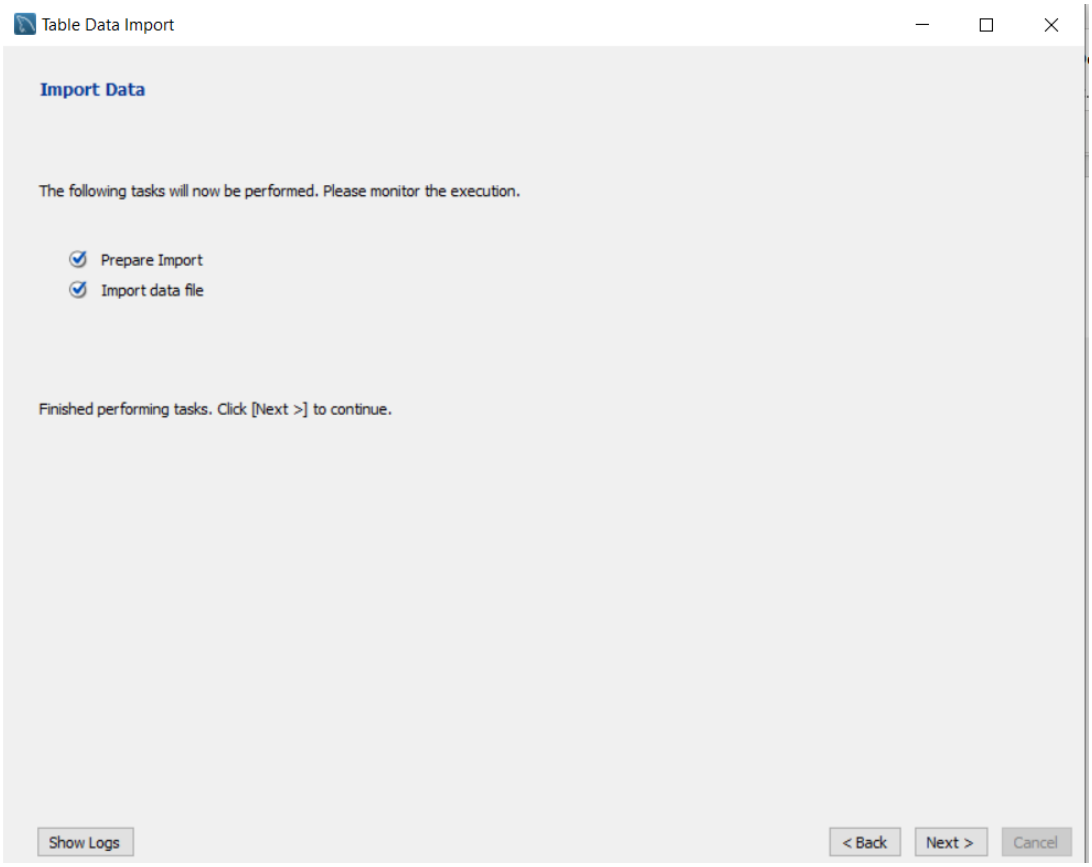
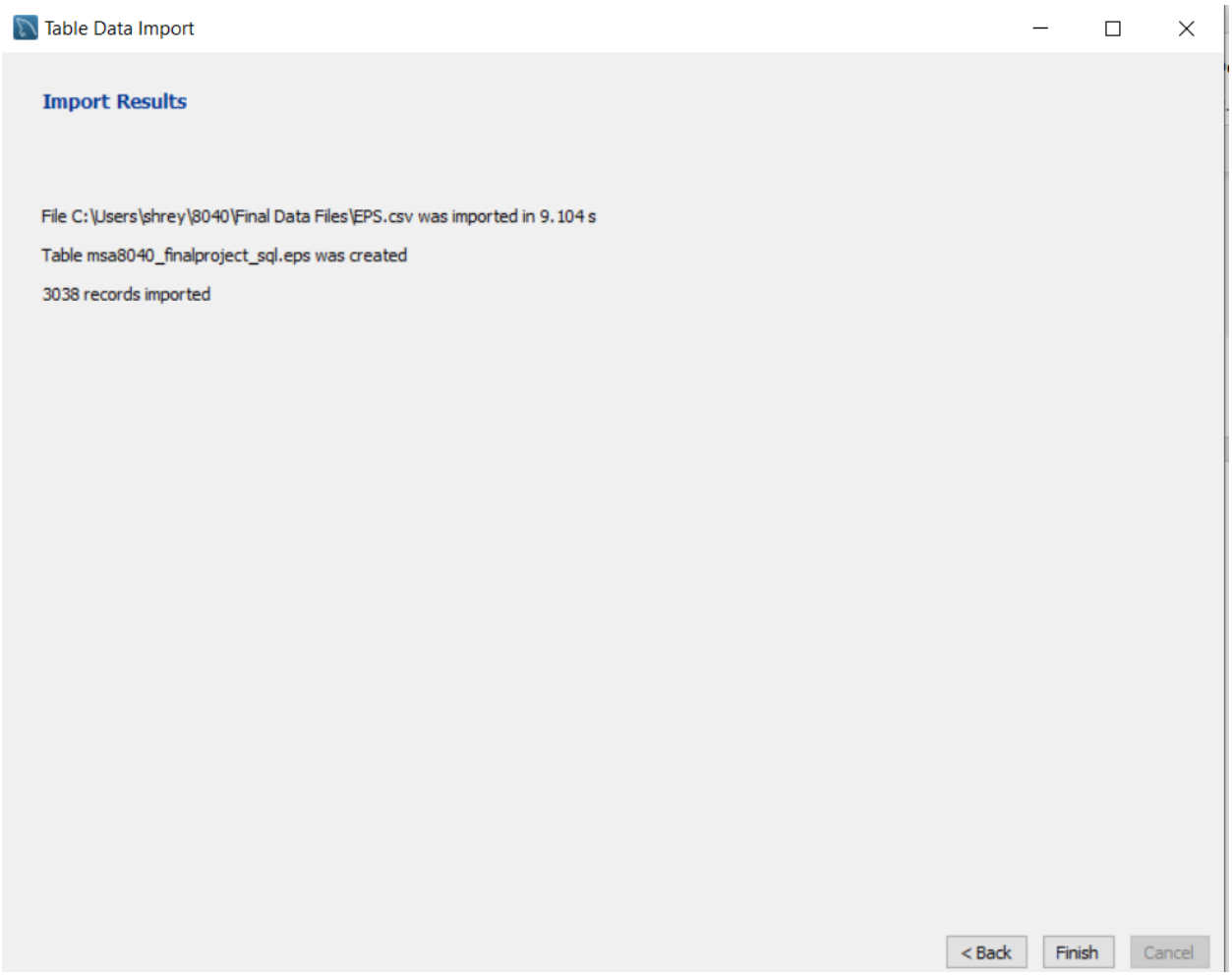
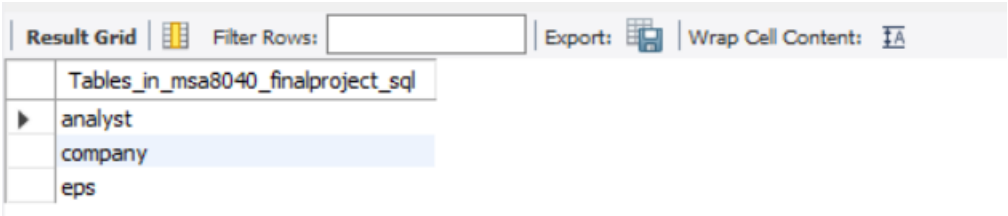


Table Data Import
Import Data
The following tasks will now be performed. Please monitor the execution.
Prepare Import
Import data file
Finished performing tasks. Click [Next >] to continue.
Show Logs
< Back
Next >
Cancel



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

```
/*Show tables in the database msa8040_finalproject_sql */
show tables;
```



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:

✓	32	18:07:52	select * from eps	3038 row(s) returned
✓	33	18:08:35	select * from company	188 row(s) returned
✓	34	18:08:44	select * from analyst	772 row(s) returned

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?

```
/* Question 1:  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? */


# Number of Analysts per company
select distinct Company_Name, count(Name) from eps group by Company_Name order by count(Name) desc;
```

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. */

```
# 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;
```


```
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?

```
/* Question 2 : Given a industry, how many companies are covered, the average number of analysts,
the average bias between the Estimize Consensus and the Reported Earnings? */

# Avg num of analysts for all the industries

select C.Company_Name, C.Industries, C.Sectors, avg(Num_analysts) as Avg_analysts
from company C
inner join eps E
where C.Company_Name = E.Company_Name
group by C.Industries;
```

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"

```
# Avg num of analysts for ' Airlines' Industries:

select E.Company_Name, C.Industries, C.Sectors, avg(Num_analysts) as Avg_analysts
from eps E
inner join company C
where Industries = ' Airlines'
and C.Company_Name = E.Company_Name;
```

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?

```
# Company having highest number of analysts with confidence score > 7

select distinct C.Company_Name, count(A.Name) as Analyst_count
from company C
inner join analyst A
where  A.Analyst_Confidence_score  =  (
    select max(A.Analyst_Confidence_score) as max_score from analyst A
    where A.Analyst_confidence_score >7
);
```

Company_Name	Analyst_count
Agilent Technologies Inc.	188

(d) Who has the largest number of followers?

```
/* Question 4: Who has the largest number of followers*/

select distinct Ticker, Company_Name, Num_Followers from company
where Num_Followers = (select max(Num_Followers) from company);
```

Ticker	Company_Name	Num_Followers
ABBV	AbbVie Inc.	650