**Stock Market Analysis**

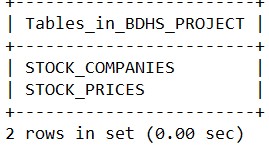
# By Anirbaan Saha

# • MySQL

**Inside the database “BDHS\_PROJECT” we have the two tables(STOCK\_COMPANIES and STOCK\_PRICES) which will be used for analysis .**

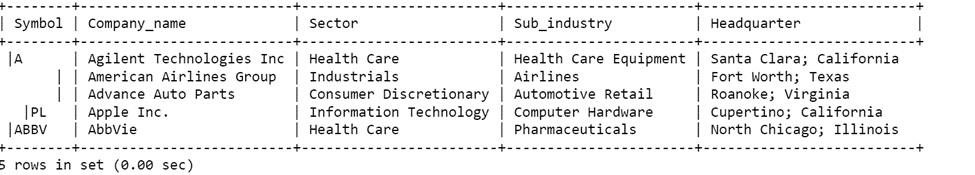
use BDHS\_PROJECT;

SHOW TABLES;

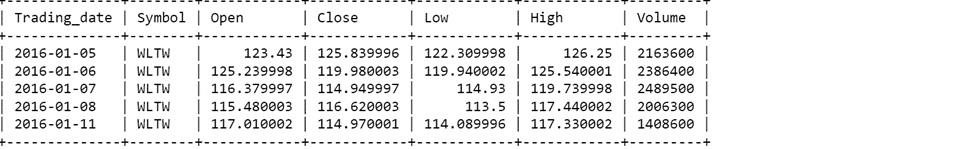


**Let us have a quick view at the datasets**

select \* from STOCK\_COMPANIES limit 5;



select \* from STOCK\_PRICES limit 5;



# • Sqoop

**We create a data pipeline using sqoop to pull the data from the MySQL server into Hive.**

sqoop import --connect jdbc:mysql://ip-10-0-1-10.ec2.internal/BDHS\_PROJECT --username labuser password simplilearn --table Stock\_companies –hive-import -hive-database stock\_db --m 1

sqoop import --connect jdbc:mysql://ip-10-0-1-10.ec2.internal/BDHS\_PROJECT --username labuser -

password simplilearn --table Stock\_prices –hive-import -hive-database stock\_db --m 1

# • Hive

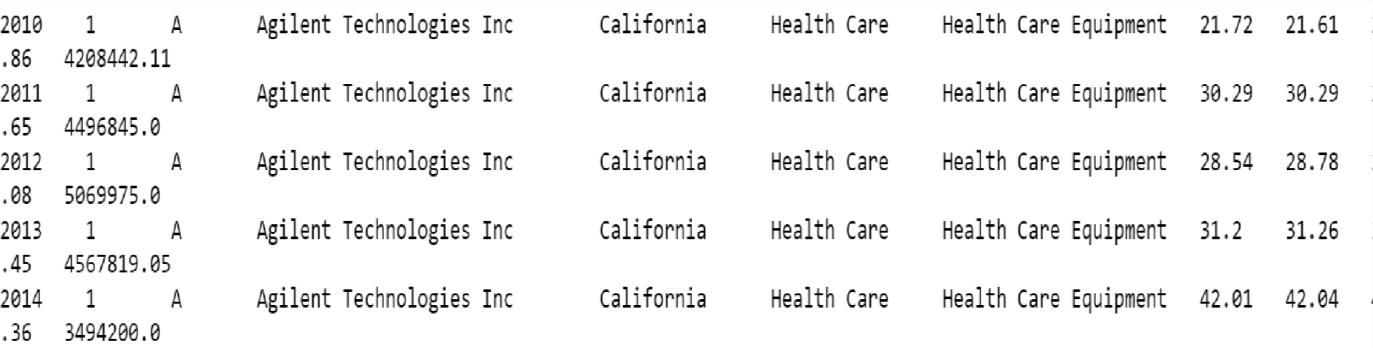
**Now we create a new hive table (stock\_data) by joining the above two hive tables (stock\_companies and stock\_prices).**

create table stock\_data as select trading\_year, trading\_month, sc.symbol,

company\_name, trim(split(headquarter,”\;”)[1]) state, sector, sub\_industry, open, close, low, high, volume from stock\_companies sc,

(select symbol, year(trading\_date) trading\_year, month(trading\_date) trading\_month, round(avg(open),2) open, round(avg(close),2) close, round(avg(low),2) low, round(avg(high),2) high, round(avg(volume),2) volume from stock\_prices group by symbol, month(trading\_date),year(trading\_date)) sp where sc.symbol=sp.symbol;

select \* from stock\_data limit 5;



# Analysis

## 1) Find the top five companies that are good for investment

create table stock\_table1 as select company\_name, min(trading\_year) min\_year,

max(trading\_year) max\_year, min(trading\_month) min\_month, max(trading\_month) max\_month from stock\_data group by company\_name;

### i)Next we find the growth percent for each company over the years

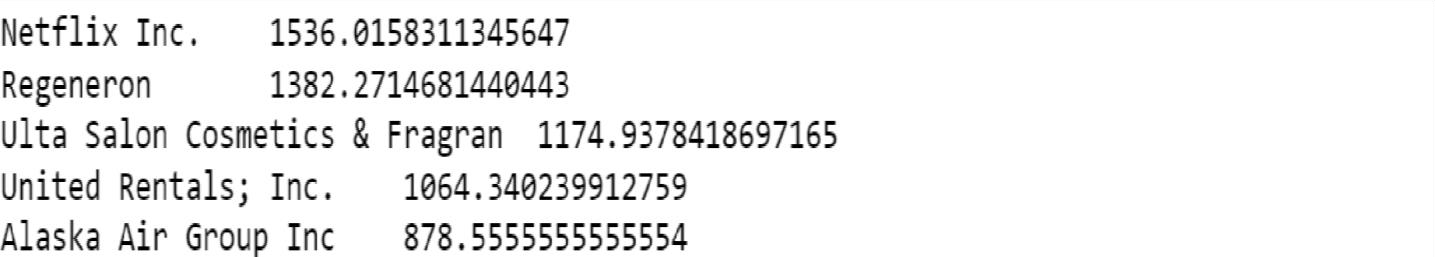
select stock\_start.company\_name, ((close-open)/open)\*100 growth\_percent from(select t1.company\_name, open from stock\_data sd, stock\_table1 t1 where sd.trading\_year = t1.min\_year and sd.trading\_month = t1.min\_month and

sd.company\_name = t1.company\_name) stock\_start, (select t1.company\_name,

close from stock\_data sd, stock\_table1 t1 where sd.trading\_year = t1.max\_year and sd.trading\_month = t1.max\_month

and sd.company\_name = t1.company\_name) stock\_end where stock\_start.company\_name = stock\_end.company\_name sort by growth\_percent desc limit 5;

### Top 5 Companies by growth



**2)Show the best-growing industry by each state, having at least two or more industries mapped.**

**i)First we calculate the growth of companies belonging to every state and capital.**

create table stock\_table2 as select state, sub\_industry, stock\_start.company\_name, ((stock\_end.close-

stock\_start.open)/stock\_start.open)\*100 growth\_percent

from (select t1.company\_name,open

from stock\_data sd, stock\_table1 t1 where sd.trading\_year=t1.min and sd.trading\_month=t1.min\_month and

sd.company\_name=t1.company\_name)stock\_start,

(select t1.company\_name, close from stock\_data sd, stock\_table1 t1 where sd.trading\_year=t1.max and sd.trading\_month=t1.max\_month and

sd.company\_name=t1.company\_name)stock\_end,

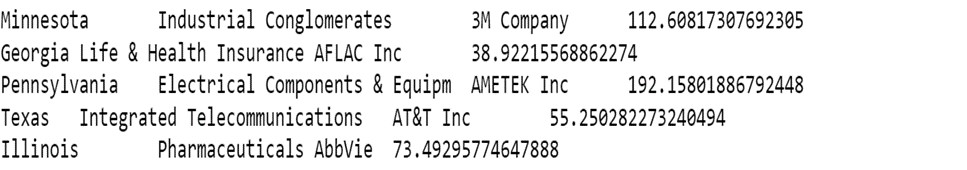
(select company\_name, state, sub\_industry

from stock\_data

group by company\_name,state,sub\_industry)sd where (stock\_end.close-stock\_start.open)>0 and

stock\_start.company\_name=stock\_end.company\_name and sd.company\_name=stock\_start.company\_name;

select \* from stock\_table2 limit 5;

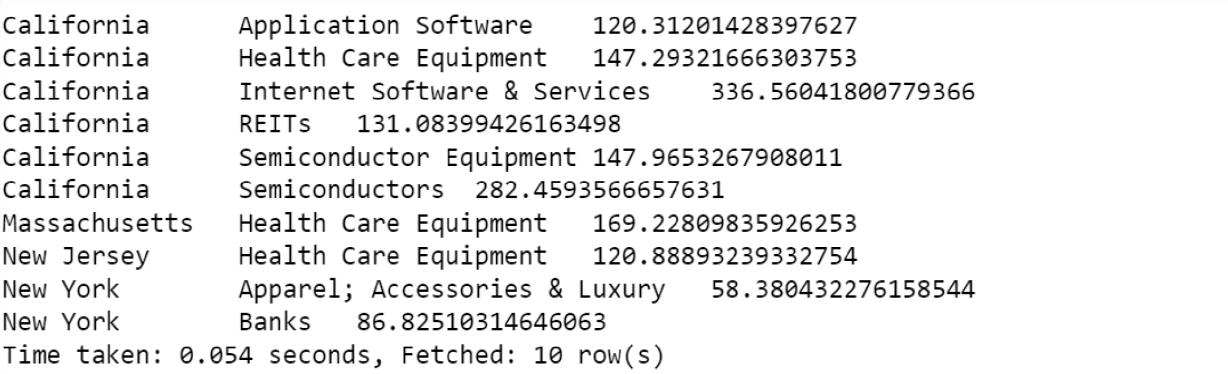


### ii)Then we group by state and capital and filter industries listed atleast twice

create table stock\_table3 as select state,sub\_industry,

avg(growth\_percent)ind\_growth from stock\_table2 group by state, sub\_industry having count(sub\_industry>=2);

Select \* from stock\_table3 limit 10;

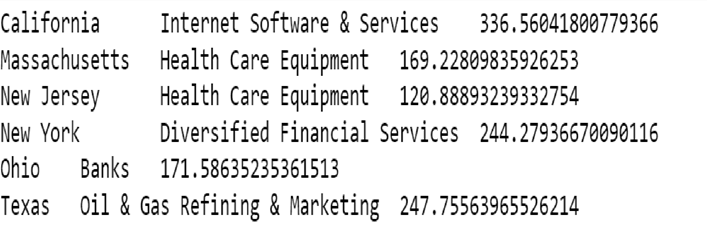


### iii)Finally we find the industry which has maximum growth by each state

select t3.state, sub\_industry, ind\_growth from stock\_table3 t3, (select state,max(ind\_growth) max\_growth from stock\_table3 group by state) max\_ind where max\_ind.state = t3.state and

t3.ind\_growth = max\_ind.max\_growth;

### best growing industry by each state



**3) For each sector find the following.**

* **Worst year**
* **Best year**

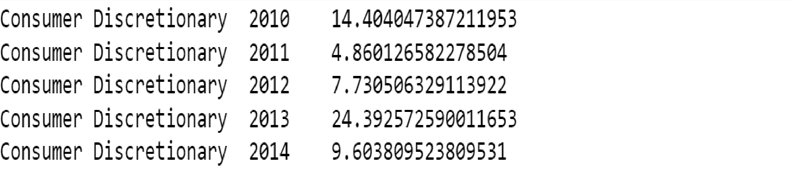
### i)AS a first step we find the growth for each sector and year

create table stock\_table4 as select open.sector, open.trading\_year, (close-open) growth from (select sector,trading\_year,avg(open) open from stock\_data where trading\_month = 1 group by sector,trading\_year) open,

(select sector,trading\_year,avg(close) close from stock\_data where trading\_month=12 group by sector,trading\_year) close

where open.sector = close.sector and open.trading\_year = close.trading\_year;

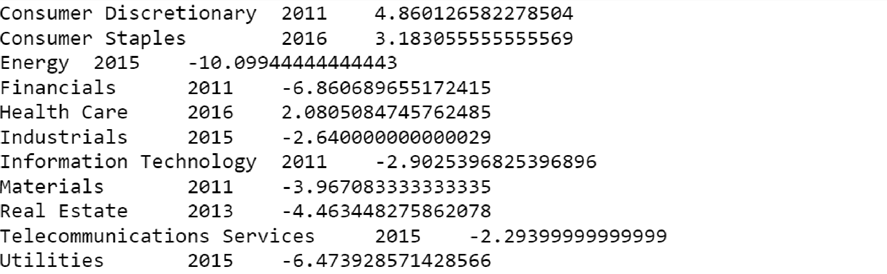
select \* from stock\_table4 limit 5;



### ii)In the year 2011, 3 sectors attained their new low

select x.sector,x.trading\_year,x.growth from stock\_table4 x, (select sector,min(growth) growth from stock\_table4 group by sector) y where x.sector=y.sector and x.growth=y.growth;

### Year in which each of the sectors were worst hit



**iii)In the year 2016 and 2014 most of the sectors enjoyed a very high growth.**

select a.sector,a.trading\_year,a.growth from stock\_table4 a,

(select b.sector,max(growth) growth from stock\_table4 group by sector) b where a.sector=b.sector and a.growth=b.growth;

### Best year for each sector

