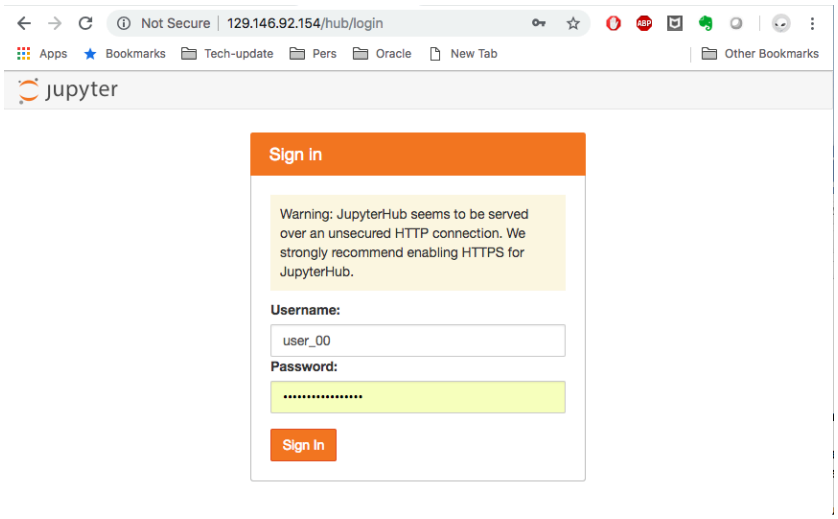
SparklineData Hands On Lab

Lab Instructions

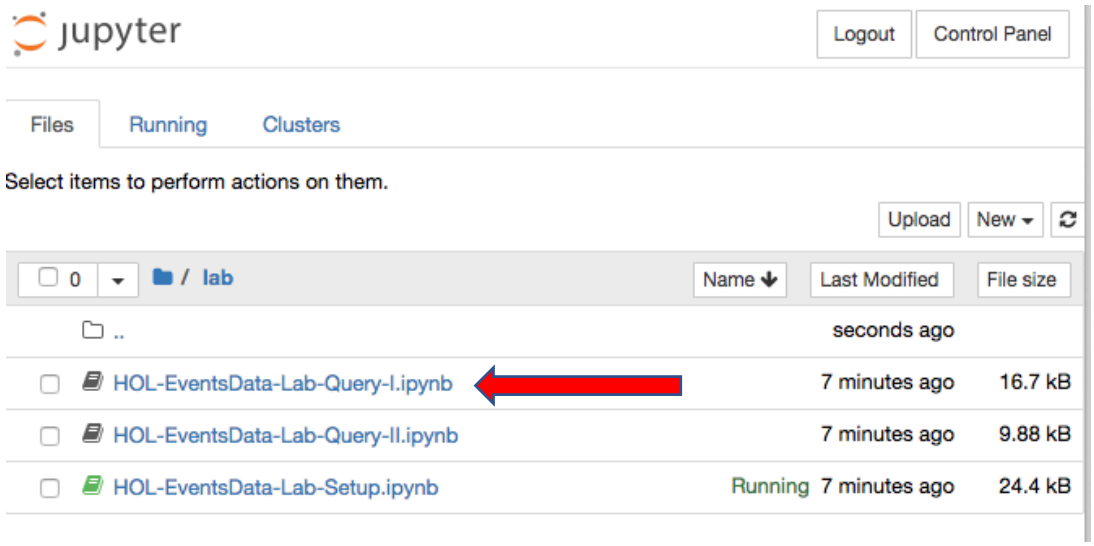
1. **Open the browser and go to the following link http://129.146.92.154/hub/login to open Jupyter notebook**



2. **You would be asked to enter your username and password, each one of you were given user id number at your table**
    1. **Username: user_<xx>**
    2. **Password: sparklinesnap2018**

3. **Once you log on to Jupyter**

    1. **Go to 'lab' directory and open the following notebook**
        1. **HOL_EventsData-Lab-Query-I**

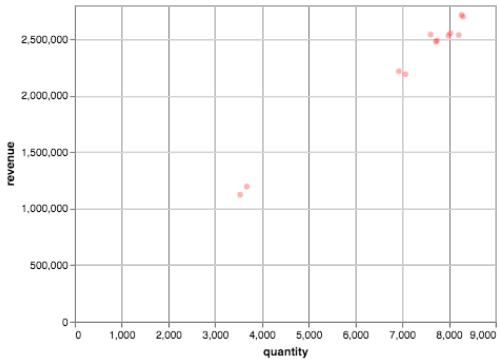


4. **Follow the instructions in the notebook and run each cell**
5. **Expected results in Query-I notebook**
    1. **Segmentation**
        1. **1.2 Query the JazzOnly segment for type, month of sales, quantity of tickets sold and the total revenue group by month order by revenue**

```
df1
```

	type	sdate	quantity	revenue
0	Jazz	MAR	8270	2715090.0
1	Jazz	OCT	8303	2698900.0
2	Jazz	SEP	8032	2552457.0
3	Jazz	APR	7609	2540940.0
4	Jazz	JUL	8210	2537106.0
5	Jazz	AUG	7989	2531728.0
6	Jazz	JUN	7743	2487415.0
7	Jazz	MAY	7720	2477804.0
8	Jazz	NOV	6929	2216240.0
9	Jazz	FEB	7070	2190557.0
10	Jazz	JAN	3678	1196203.0
11	Jazz	DEC	3536	1124268.0

```
time: 14.2 ms
```



2. **1.3 Query the SportsOnly segment for type, month of sales, quantity of tickets sold and the total revenue group by month order by revenue**

```
df2
```

	type	sdate	quantity	revenue
0	Sports	MAR	5983	1964650.0
1	Sports	MAY	6042	1938157.0
2	Sports	AUG	5689	1884797.0
3	Sports	JUL	5995	1883230.0
4	Sports	OCT	5923	1881128.0
5	Sports	SEP	5816	1823764.0
6	Sports	JUN	5476	1793375.0
7	Sports	APR	5404	1763644.0
8	Sports	FEB	5124	1601135.0
9	Sports	NOV	5054	1580565.0
10	Sports	JAN	3122	971223.0
11	Sports	DEC	2567	796433.0

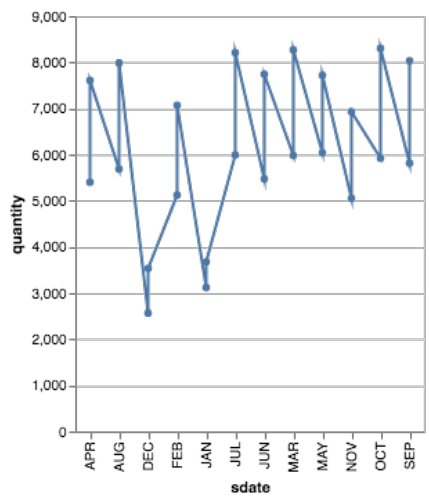
```
time: 20.2 ms
```

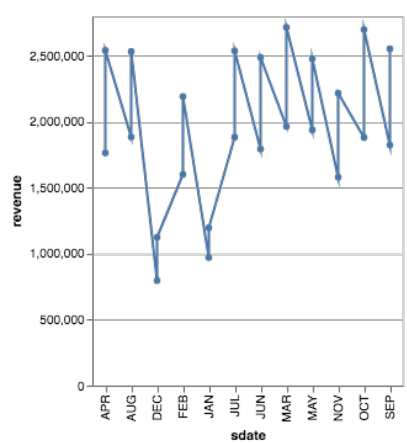
3. **1.4 Combine both the results and compare revenue and quantity them each month**

	type	sdate	quantity	revenue
0	Jazz	MAR	8270	2715090.0
1	Jazz	OCT	8303	2698900.0
2	Jazz	SEP	8032	2552457.0
3	Jazz	APR	7609	2540940.0
4	Jazz	JUL	8210	2537106.0
5	Jazz	AUG	7989	2531728.0
6	Jazz	JUN	7743	2487415.0
7	Jazz	MAY	7720	2477804.0
8	Jazz	NOV	6929	2216240.0
9	Jazz	FEB	7070	2190557.0
10	Jazz	JAN	3678	1196203.0
11	Jazz	DEC	3536	1124268.0

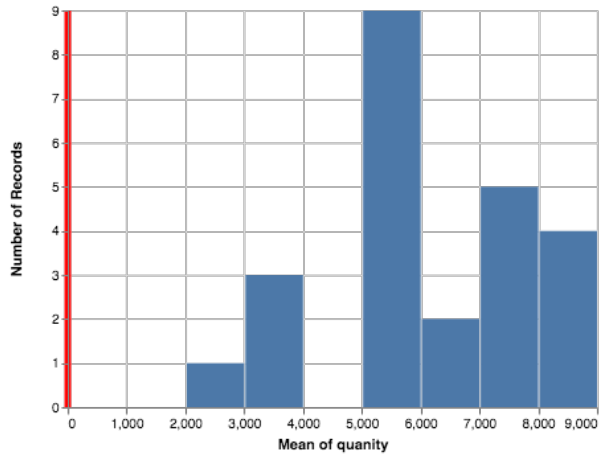
	type	sdate	quantity	revenue
0	Sports	MAR	5983	1964650.0
1	Sports	MAY	6042	1938157.0
2	Sports	AUG	5689	1884797.0
3	Sports	JUL	5995	1883230.0
4	Sports	OCT	5923	1881128.0
5	Sports	SEP	5816	1823764.0
6	Sports	JUN	5476	1793375.0
7	Sports	APR	5404	1763644.0
8	Sports	FEB	5124	1601135.0
9	Sports	NOV	5054	1580565.0
10	Sports	JAN	3122	971223.0
11	Sports	DEC	2567	796433.0

time: 57.9 ms





4. **1.5 Draw a histogram**



2. **Repeat Customer Analysis**
   1. **Creates a view with the first ticket sales time and the most recent user activity date for each customer**

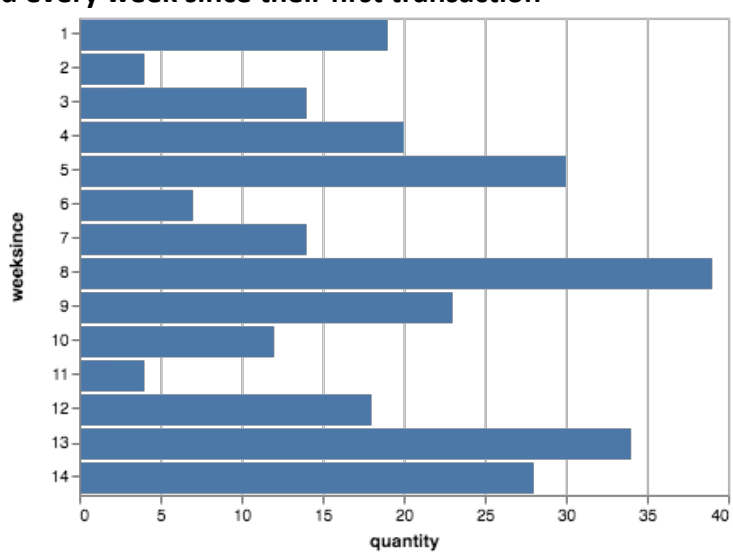
	start	weeksince	quantity	price	dist_count
0	1	1	144	46602.0	68
1	1	4	39	16646.0	19
2	1	5	138	44240.0	62
3	1	6	47	15612.0	23
4	1	8	22	5159.0	9
5	1	9	146	42099.0	72
6	1	10	52	11083.0	22
7	1	13	116	31361.0	58
8	1	14	147	42054.0	66
9	2	3	52	15406.0	23
10	2	4	140	39377.0	67
11	2	5	59	18978.0	27
12	2	7	23	6213.0	11
13	2	8	193	53590.0	87
14	2	9	81	25746.0	40
15	2	12	125	43856.0	59
16	2	13	156	52348.0	73
17	5	1	47	15041.0	24
18	5	2	24	5790.0	10
19	5	4	12	5094.0	6
20	5	5	48	16395.0	21
21	5	6	24	6786.0	12
22	5	9	42	15567.0	24
23	5	10	60	19698.0	27

   2. **For each Jazzonly user get the amount of tickets purchased and the price paid every week since their first transaction**

```
df2
```

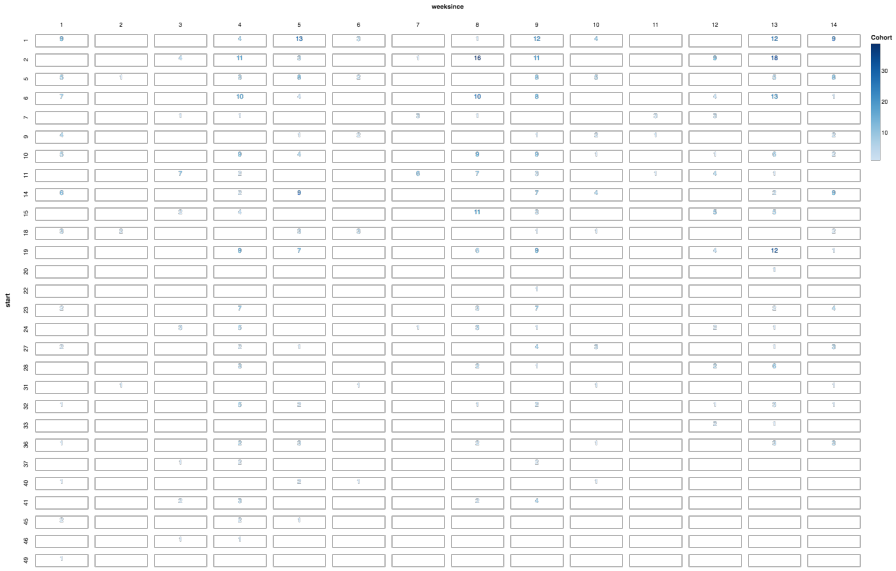
	start	weeksince	quantity	price	dist_count
0	1	1	19	7505.0	9
1	1	4	10	3112.0	4
2	1	5	27	14529.0	13
3	1	6	7	475.0	3
4	1	8	2	284.0	1
5	1	9	23	6736.0	12
6	1	10	12	2091.0	4
7	1	13	25	7158.0	12
8	1	14	28	7199.0	9
9	2	3	7	3010.0	4
10	2	4	20	3998.0	11
11	2	5	7	864.0	3
12	2	7	1	391.0	1
13	2	8	39	9746.0	16
14	2	9	22	6275.0	11
15	2	12	18	5663.0	9
16	2	13	34	6419.0	18
17	5	1	12	3315.0	5
18	5	2	1	204.0	1
19	5	4	8	1751.0	3
20	5	5	15	3108.0	8
21	5	6	6	1594.0	2

3. **For each SportsOnly user get the amount of tickets purchased and the price paid every week since their first transaction**

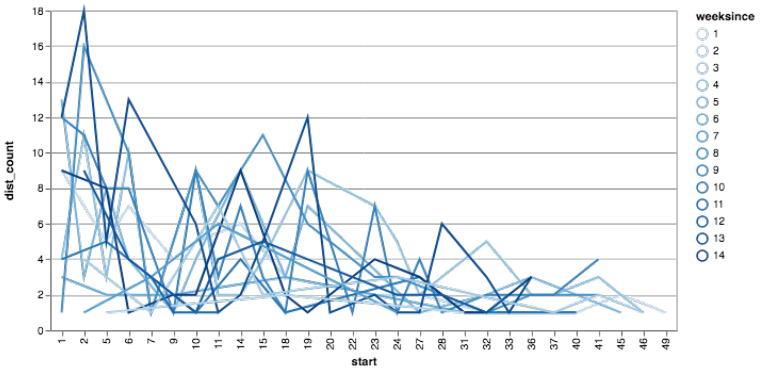


3. **Cohort analysis**

1. **Draw a chart showing the behavior of Cohorts who bought tickets together and their subsequent behavior every week**

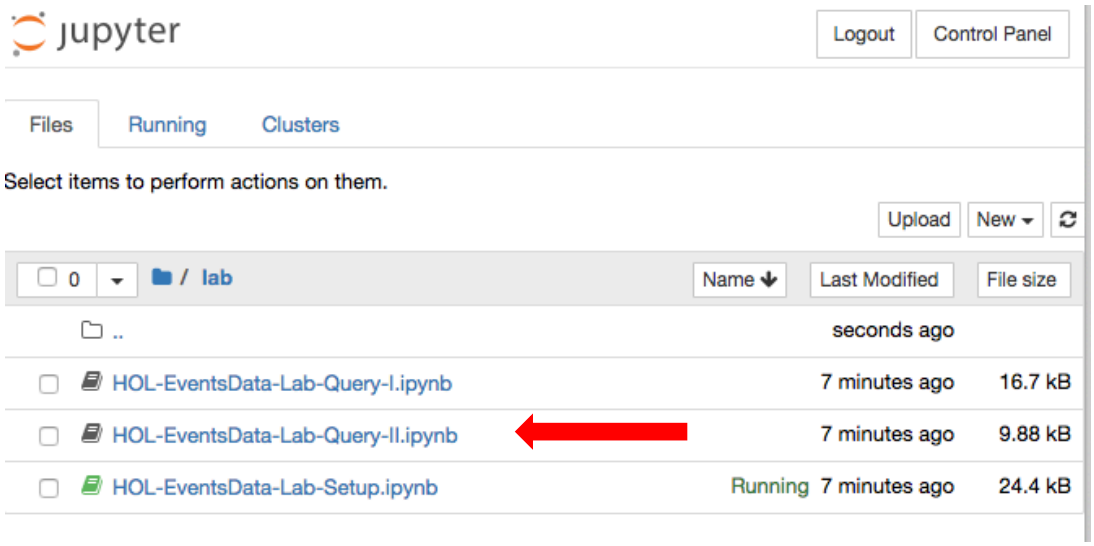


2. **Returning sports customers - customers who are coming repeatedly**



HOL_EventsData-Lab-Query-II

1. Open the second notebook HOL_EventsData-Lab-Query-II



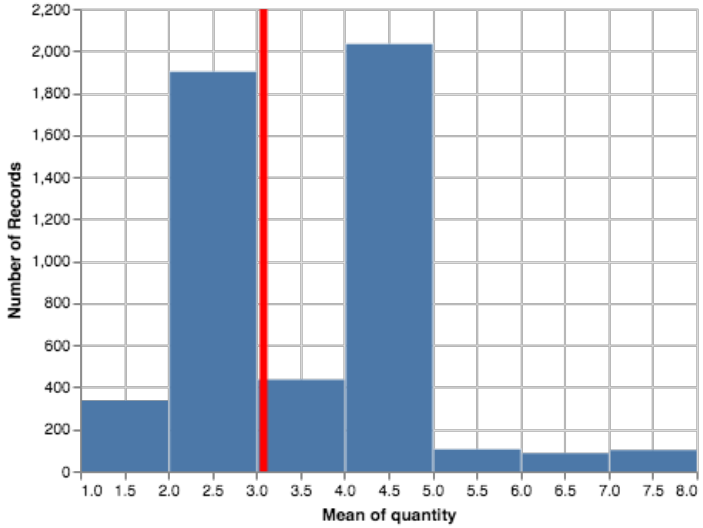
2. Follow the instructions in the notebook to run one cell at a time

1. Query-1 on quantity of tickets sold and revenue/cost by date:
    - Compare sales of all users to users who liked Jazz and Concerts

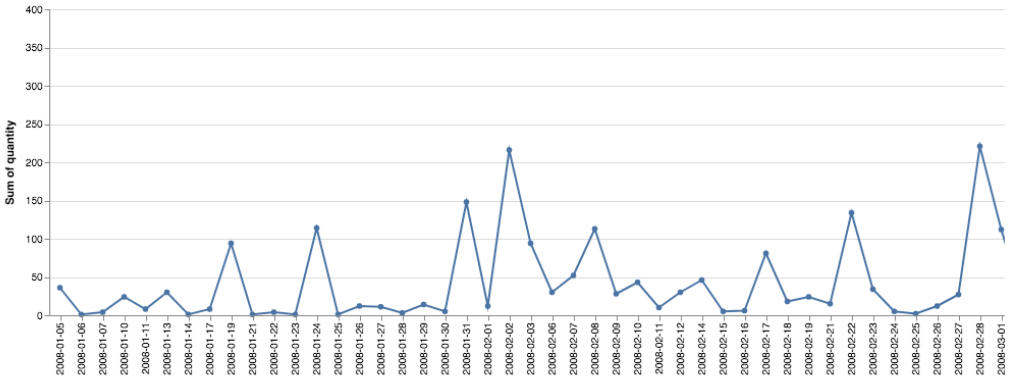
	adate	users_buyer_city	quantity	price	quantity_ratio	price_ratio
0	2008-01-31	Webster Groves	4	9900.0	400.0	49500.0
1	2008-08-10	Sedalia	4	9772.0	400.0	48860.0
2	2008-05-06	Salt Lake City	4	9456.0	400.0	47280.0
3	2008-10-10	Murrieta	4	9432.0	400.0	47160.0
4	2008-09-09	Monroe	4	9392.0	400.0	46960.0
5	2008-04-01	Santa Rosa	4	9844.0	400.0	46876.0
6	2008-06-29	Bayamon	8	8892.0	800.0	44460.0
7	2008-09-09	Monroe	4	8832.0	400.0	44160.0
8	2008-01-24	Fort Collins	4	9636.0	400.0	43800.0
9	2008-04-30	Mequon	4	8752.0	400.0	43760.0
10	2008-07-29	Janesville	4	9112.0	400.0	43390.0
11	2008-06-26	Broken Arrow	4	9028.0	400.0	42990.0
12	2008-05-30	Warren	4	8994.0	400.0	42829.0
13	2008-08-10	Livonia	4	9772.0	400.0	42487.0
14	2008-04-27	San Mateo	4	9724.0	400.0	42278.0
15	2008-03-23	Laguna Hills	4	9624.0	400.0	41843.0
16	2008-06-24	Half Moon Bay	4	8716.0	400.0	41505.0
17	2008-08-10	Texas City	4	9772.0	400.0	40717.0
18	2008-03-23	Wisconsin Dells	4	9624.0	400.0	40100.0

- Draw
    - Histogram of number of tickets sold

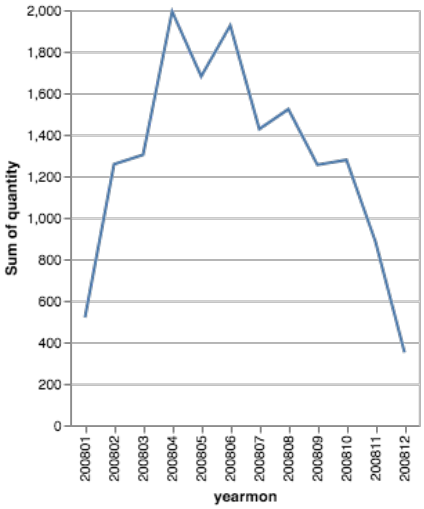
1- Draw histogram of number of tickets sold



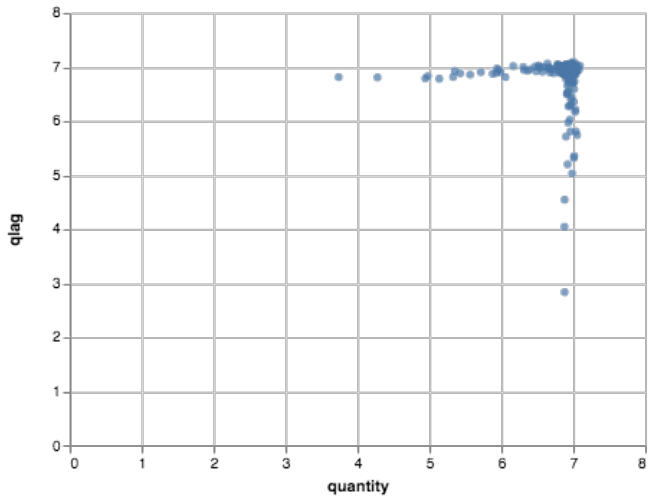
- Time series of tickets sold – daily



- Plot Quantity of tickets sold per month



2. Query-2 Find quantity sold compared to quantity sold over a 40 day window
    - Scatter plot of quantity of tickets sold vs quantity of tickets sold 40 days ago



- Auto correlation plot