

# class2



default ▼

```
%spark
import org.apache.spark.sql.functions._
import org.joda.time.format.DateTimeFormat

import org.apache.spark.sql.functions._
import org.joda.time.format.DateTimeFormat
```

Took 27 sec. Last updated by anonymous at January 20 2017, 11:29:29 AM.

FINISHED    

```
val inputPath = "/Users/jyothi/Downloads/data/*.csv"
val airTraffic = sqlContext.read
    .format("com.databricks.spark.csv")
    .option("header", "true") // Use first line of all files as header
    .option("delimiter", ",")
    .option("inferSchema", "true") // Automatically infer data types
    .load(inputPath)
```

```
inputPath: String = /Users/jyothi/Downloads/data/*.csv
airTraffic: org.apache.spark.sql.DataFrame = [Year: int, Month: int ... 27 more fields]
```

Took 3 min 28 sec. Last updated by anonymous at January 20 2017, 11:36:15 AM.

FINISHED ▶ ⌵ ⌶ ⚙

```
val calcDayOfYear = udf(  
  (dayOfMonth: Int, month: Int, year: Int) => {  
    val dateFormat = DateTimeFormat.forPattern("dd/MM/yyyy")  
    dateFormat.parseDateTime(s"$dayOfMonth/$month/$year").getDayOfYear()  
  }  
)
```

```
warning: Class org.joda.convert.FromString not found - continuing with a stub.
warning: Class org.joda.convert.ToString not found - continuing with a stub.
warning: Class org.joda.convert.ToString not found - continuing with a stub.
calcDayOfYear: org.apache.spark.sql.expressions.UserDefinedFunction = UserDefinedFunction(<function3>,IntegerType,Some(List(IntegerType, IntegerType, IntegerType)))
```

Took 4 sec. Last updated by anonymous at January 20 2017, 11:37:01 AM.

FINISHED ▶ ↻ 📖 ⚙️

```
val calcRoute = udf(
  (origin: String, dest: String) => s"$origin - $dest"
)
val calcHourOfArrival = udf(
  (arrTime: String) => arrTime.slice(0,arrTime.size-2)
)
```

```
calcRoute: org.apache.spark.sql.expressions.UserDefinedFunction = UserDefinedFunction(<function2>,StringType,Some(List(StringType, StringType)))
calcHourOfArrival: org.apache.spark.sql.expressions.UserDefinedFunction = UserDefinedFunction
(<function1>,StringType,Some(List(StringType)))
```

Took 4 sec. Last updated by anonymous at January 20 2017, 11:37:07 AM.

Zeppelin

val featuredTraffic = airTraffic

class2

withColumn("DayOfYear", calcDayOfYear(airTraffic("DayOfMonth"), airTraffic("Month"), airTraffic("Year")))

withColumn("HourOfArr", calcHourOfArrival(airTraffic("ArrTime")))

withColumn("Route", calcRoute(airTraffic("Origin"), airTraffic("Dest")))

FINISHED

▶

⌵

📖

⚙️

default

featuredTraffic: org.apache.spark.sql.DataFrame = [Year: int, Month: int ... 30 more fields]

Took 2 sec. Last updated by anonymous at January 20 2017, 11:37:10 AM.

featuredTraffic.registerTempTable("air\_traffic")

FINISHED

▶

⌵

📖

⚙️

warning: there was one deprecation warning; re-run with -deprecation for details

Took 3 sec. Last updated by anonymous at January 20 2017, 11:37:17 AM.

%sql

select DayOfYear, count(\*) as NrOfFlights, avg(DepDelay) as AvgDepDelay, avg(ArrDelay) as AvgArrDelay from air\_traffic where DayOfYear having NrOfFlights > 100000

FINISHED

▶

⌵

📖

⚙️

📊

📈

📉

📊

📈

📉

⬇️

⬆️

DayOfYear	NrOfFlights	AvgDepDelay
148	197,353	7.59018
243	201,007	6.73581
31	200,810	8.38126
85	204,282	7.78211
137	203,693	8.61366
251	199,315	4.20382
65	206,041	9.45782
53	204,590	11.6728
255	201,085	4.60038

Took 3 min 6 sec. Last updated by anonymous at January 20 2017, 10:39:35 AM.

%sql

select HourOfArr, count(\*) as NrFlights, avg(ArrDelay) as AvgArrDelay from air\_traffic where 1(Wednesday)14(Thursday)15(Friday)16(Saturday)17(Sunday)} and Origin = "\${org=ATL,ATL|PHX|P:

FINISHED

▶

⌵

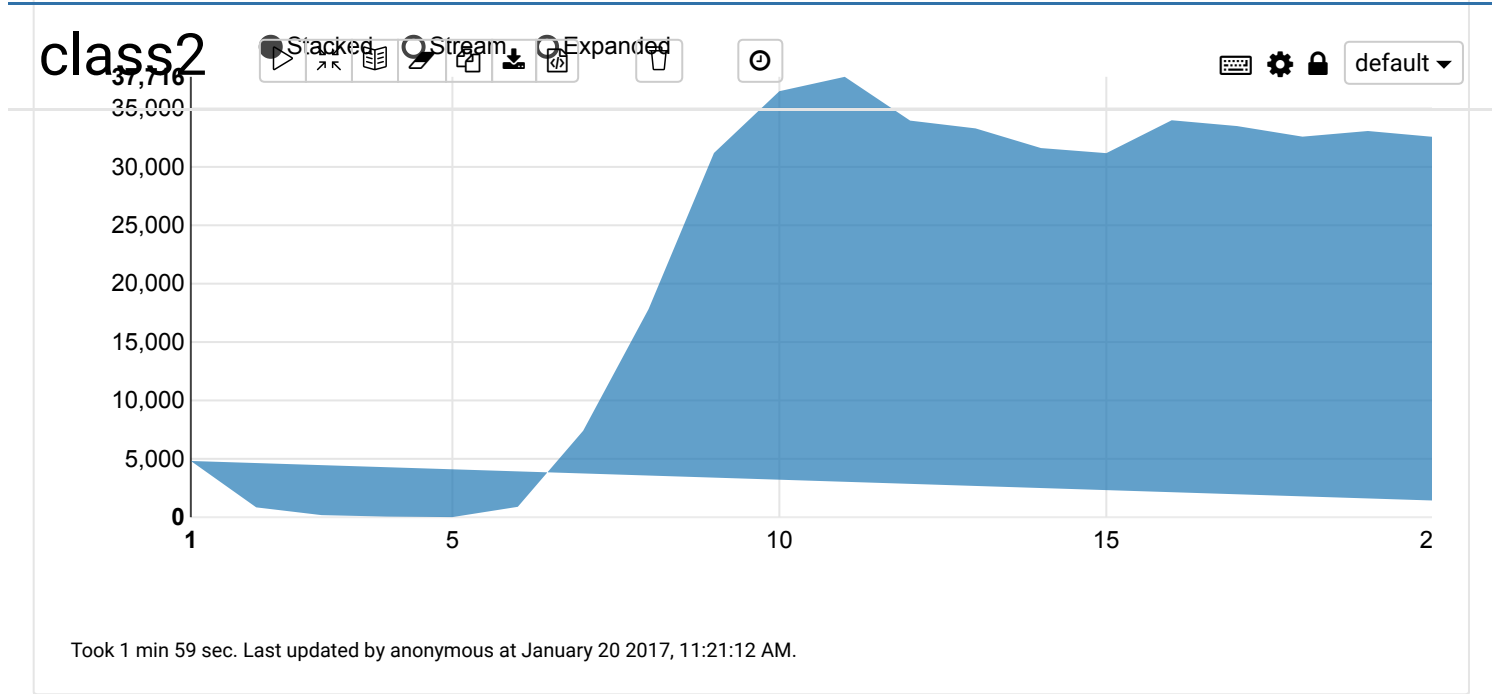
📖

⚙️

org

ATL

day



```
%pyspark
```

FINISHED ▶ ⌵ ⌵ ⌵ ⌵

Took 17 sec. Last updated by anonymous at January 20 2017, 6:44:44 AM. (outdated)

```
%pyspark
import seaborn as sns
```

FINISHED ▶ ⌵ ⌵ ⌵ ⌵

Took 10 sec. Last updated by anonymous at January 20 2017, 11:37:41 AM.

```
%pyspark

import matplotlib
matplotlib.use('Agg')
import matplotlib.pyplot as plt
```

/Users/jyothi/anaconda/lib/python3.5/site-packages/matplotlib/\_\_init\_\_.py:1357: UserWarning:  
This call to matplotlib.use() has no effect  
because the backend has already been chosen;  
matplotlib.use() must be called \*before\* pylab, matplotlib.pyplot,  
or matplotlib.backends is imported for the first time.  
warnings.warn(\_use\_error\_msg)

Took 0 sec. Last updated by anonymous at January 20 2017, 11:38:11 AM.

```
%pyspark
from io import StringIO
```

FINISHED ▶ ⌵ ⌵ ⌵ ⌵

blanco

http://localhost:8080/#/notebook/2C9P2XC6K

class2



FINISHED



default ▾

```
%pyspark
def show(p):
    img = StringIO.StringIO()
    p.savefig(img, format='svg')
    img.seek(0)
    print( "%html " + img.buf)
df = sqlContext.sql("SELECT Dest, Month, count(*) as NrOfFlights, avg(ArrDelay) as AvgArrDelay
Month")
data = df.toPandas()
```

Took 2 min 14 sec. Last updated by anonymous at January 20 2017, 11:41:07 AM.

%pyspark

FINISHED



```
value = "AvgArrDelay"
x = "Dest"
grouping = ["Month"]

heatmap_data = data.pivot_table(values=value, index=x, columns=grouping)
heatmap_data = heatmap_data[0:100]
```

Took 0 sec. Last updated by anonymous at January 20 2017, 11:53:52 AM.

```
%pyspark
a4_dims = (len(heatmap_data.columns),50)
fig, ax = plt.subplots(figsize=a4_dims)
ax.set_title("Avg Arrival Delay")
sns.heatmap(heatmap_data, ax=ax, annot=True, fmt=".02f")

<matplotlib.axes._subplots.AxesSubplot object at 0x11400ef60>
```

FINISHED



Took 2 sec. Last updated by anonymous at January 20 2017, 11:54:02 AM.

```
%pyspark
def show(p):
    img = StringIO()
    p.savefig(img, format='svg')
    img.seek(0)
    print( "%html " + img.read())
```

FINISHED




Took 0 sec. Last updated by anonymous at January 20 2017, 11:54:09 AM.

```
%pyspark
show(plt)
```

FINISHED



Avg Arrival Delay												
ABE	9.70	7.74	7.60	5.87	7.00	12.95	11.98	9.34	3.13	3.38	4.81	11.19
ABI	7.11	7.28	6.39	4.82	5.68	14.84	11.49	10.33	2.86	4.80	6.85	11.83
ABQ	6.84	7.13	6.75	5.28	5.52	9.62	8.21	7.24	2.66	5.37	4.45	10.74
ABY	8.73	11.08	6.15	1.60	4.22	12.88	14.79	14.37	6.63	8.29	6.79	11.58
ACK						3.56	26.33	24.23	19.51	2.89		
ACT	3.65	3.60	2.07	0.39	0.67	6.58	4.22	5.74	0.61	1.49	2.88	2.43
ACV	13.04	10.16	4.58	2.46	6.08	8.77	6.40	7.67	4.62	9.13	7.81	18.28
ACY	3.78	6.03	2.58	3.40	0.45	14.61	18.11	13.02	4.03	4.41	4.21	5.87
ADK	23.03	10.35	16.67	13.34	4.10	-0.30	-2.83	7.89	12.53	17.65	9.49	4.59
ADQ	11.33	8.33	7.97	4.14	5.36	9.47	10.02	8.92	7.18	6.29	7.00	13.50
AEX	5.38	8.01	5.69	4.13	3.95	14.44	12.21	10.85	3.05	3.89	5.80	9.76
AGS	7.03	9.39	8.91	8.43	7.71	15.18	17.69	13.38	6.43	6.13	6.71	9.33
AKN	12.08	16.00	16.82	16.90	8.45	14.93	8.55	10.69	10.21	13.36	6.17	14.37
ALB	9.72	8.77	9.34	7.01	8.16	13.95	14.19	12.24	3.89	4.61	5.24	11.27
ALO	-0.52	4.08	1.38	2.96	-4.20	7.27	5.55	5.94	-3.33	0.65	6.28	15.09
AMA	6.71	8.21	8.63	6.39	8.60	13.72	9.83	8.63	4.54	6.65	6.35	11.61
ANC	11.92	9.06	10.08	4.88	8.11	11.41	11.90	12.86	7.88	7.64	9.79	17.24
ANI	19.00	13.06	9.90	19.64	10.52	14.17	30.29	9.33	1.00	10.00	13.46	8.58
APF	9.60	10.02	12.19	5.77	4.12	15.13	27.15	22.19	4.67	3.13	9.16	8.81
ASE	18.70	16.16	10.50	1.69	-0.06	5.41	6.18	6.93	1.93	0.19	-3.16	23.05
ATL	8.44	9.97	8.65	5.18	5.40	11.75	13.32	10.03	4.41	5.62	6.72	12.63
ATW	14.09	16.72	10.83	7.87	4.33	12.32	10.57	9.39	3.33	3.52	6.33	20.55
AUS	6.63	6.61	7.03	5.03	6.01	10.38	9.04	7.36	2.08	4.43	4.71	9.68
AVL	5.64	5.62	4.95	4.23	4.11	12.32	13.23	12.01	4.24	6.21	4.24	9.57
AVP	13.64	10.37	8.93	7.91	8.86	16.60	14.68	12.03	4.61	3.93	5.26	15.07
AZO	13.04	10.85	8.52	5.83	5.62	9.87	7.97	6.13	1.93	2.15	5.38	15.28
BDL	7.44	7.16	7.91	5.65	6.35	12.58	11.54	9.95	2.84	3.19	3.74	8.69
BET	10.43	7.79	7.84	5.67	6.32	8.76	10.58	14.15	6.35	8.32	11.86	13.64
BFF												
BFL	7.39	5.20	3.80	1.29	2.16	4.41	3.56	2.47	1.76	3.92	2.83	10.91
3GM	10.13	8.25	4.78	4.77	5.75	11.42	7.87	7.11	2.25	2.10	4.06	11.62

30

http://localhost:8080/#/notebook/2C9P2XC6K

5/13

class2	BIL	9.55	7.62	6.04	3.21	4.89	9.30	9.58	7.63	2.29	3.30	3.86	15.24								
	BIS	7.39	8.31	7.27	2.75	4.65	10.94	6.97	6.82	0.53	1.37	2.86	14.01								
	BIJ						2.80	0.91	4.14	9.46	2.00	-7.00									
	BLI	8.10	3.14	4.29	-6.49	-2.17	2.26	4.08	11.05	7.79	7.18	9.58	20.38								
	BMI	16.57	17.23	13.92	7.20	6.23	14.18	14.72	10.89	5.27	4.06	9.08	17.52								
	BNA	6.17	5.68	5.73	4.15	5.22	9.73	9.02	6.68	1.19	3.08	3.41	8.88								
	BOI	10.99	8.98	7.10	4.59	5.36	9.04	8.84	8.55	3.66	5.77	6.02	16.23								
	BOS	8.04	7.18	8.88	6.75	7.13	13.57	12.13	11.28	4.32	4.54	4.86	9.58								
	BPT	1.13	2.85	2.28	0.94	0.67	9.87	5.02	3.46	0.42	-0.46	4.71	1.36								
	BQK	4.84	9.88	9.98	5.50	8.91	16.09	17.60	18.34	8.30	8.40	6.95	11.60								
	BQN	8.44	10.49	13.20	12.39	11.62	20.15	19.66	15.73	4.15	0.77	5.93	18.07								
	BRO	5.43	6.48	6.23	3.24	4.32	10.12	6.94	4.09	-1.81	1.28	4.76	9.28								
	BRW	8.16	10.43	9.52	2.54	7.40	10.68	14.19	13.22	5.01	4.69	4.78	11.41								
	BTM	8.84	4.50	3.20	1.90	1.40	4.59	0.99	2.08	-2.66	-0.80	1.57	15.26								
	BTR	6.68	8.84	7.99	6.46	5.93	12.90	11.85	9.01	4.30	4.66	6.52	9.80								
	BTV	11.39	11.13	12.97	8.74	8.59	15.05	16.72	14.68	3.51	4.31	4.88	14.34								
Dest	BUF	10.00	9.66	9.04	8.59	7.14	14.00	13.76	11.37	3.49	4.48	5.41	12.65								
	BUR	8.74	9.70	6.69	6.14	4.98	7.59	7.19	7.61	4.29	7.29	7.52	12.33								
	BWI	6.55	6.00	6.79	4.55	6.07	12.24	11.74	9.68	2.19	2.99	4.15	8.37								
	BZN	13.86	10.58	6.19	1.19	3.73	8.36	6.82	6.09	1.31	0.76	2.59	17.48								
	CAE	10.67	11.28	9.58	7.46	7.41	13.74	14.34	11.75	4.84	4.97	6.88	13.77								
	CAK	11.06	12.23	8.02	5.25	5.37	11.24	13.29	10.96	2.22	4.28	6.18	14.63								
	CBM																				
	CCR											10.67	10.37	13.46							
	CDC	7.62	5.55	1.39	-0.30	0.57	1.34	2.66	0.59	1.39	0.24	2.64	6.99								
	CDV	12.15	10.34	7.67	9.68	12.24	19.11	17.79	21.97	15.52	9.68	5.97	17.12								
	CEC	12.54	8.11	5.32	1.39	2.13	5.27	5.36	7.55	3.14	4.75	7.13	14.39								
	CHA	8.30	8.53	6.39	5.86	5.83	13.02	12.86	8.81	3.88	4.15	5.75	10.76								
	CHO	8.87	8.31	3.54	3.27	3.99	8.88	10.49	9.74	2.43	3.85	4.24	5.19								
	CHS	7.47	8.86	8.90	6.97	6.52	11.65	12.41	9.29	4.04	4.97	4.90	11.27								
	CIC	13.89	11.62	3.07	2.11	8.35	10.49	11.07	6.59	5.68	8.94	12.96	16.57								
	CID	12.33	10.10	8.71	6.34	8.32	12.54	10.12	8.67	2.61	2.73	4.87	14.52								
	CKB	7.83																			
	CLD	3.40	2.67	1.39	-1.40	-0.36	-0.26	-0.42	-2.08	-0.90	1.40	-0.51	0.71								
	CLE	8.54	6.45	6.69	4.66	5.31	10.52	9.65	8.08	2.01	3.10	4.30	10.53								
	CLL	4.25	4.44	2.29	0.95	1.57	7.62	4.02	5.82	-1.01	0.66	2.21	2.78								
	CLT	5.79	5.16	5.51	3.94	3.53	8.75	7.98	6.23	0.78	1.68	2.58	7.47								
	CMH	9.85	8.54	8.68	6.86	7.83	13.02	12.12	10.18	3.64	4.36	5.53	12.03								
	CMI	25.22	17.63	14.86	10.15	11.91	15.88	15.69	12.73	4.01	4.44	10.23	22.05								
	CMX	12.46	11.20	-6.45	10.49	1.79	5.06	24.55	13.74	12.11	-0.40	6.33	34.69								
	COD	11.85	12.70	2.41	-0.49	1.07	2.88	4.74	3.10	-0.52	4.09	1.35	16.36								

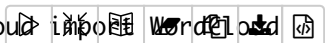
# class2

A screenshot of a web browser interface. On the left is a solid blue vertical sidebar. Within this sidebar, there are three white icons: a keyboard, a gear (settings), and a padlock. To the right of the sidebar is a white area containing a dropdown menu. The dropdown menu is open, showing the text 'default' followed by a downward-pointing arrow. Below the dropdown menu, the text '-30' is visible.

FINISHED ▶ 🔍 📖 ⚙️

class2Untitled1Untitled2Untitled3Untitled4Untitled5Untitled6Untitled7Untitled8Untitled9Untitled10Untitled11Untitled12Untitled13Untitled14Untitled15Untitled16Untitled17Untitled18Untitled19Untitled20Untitled21Untitled22Untitled23Untitled24Untitled25Untitled26Untitled27Untitled28Untitled29Untitled30Untitled31Untitled32Untitled33Untitled34Untitled35Untitled36Untitled37Untitled38Untitled39Untitled40Untitled41Untitled42Untitled43Untitled44Untitled45Untitled46Untitled47Untitled48Untitled49Untitled50Untitled51Untitled52Untitled53Untitled54Untitled55Untitled56Untitled57Untitled58Untitled59Untitled60Untitled61Untitled62Untitled63Untitled64Untitled65Untitled66Untitled67Untitled68Untitled69Untitled70Untitled71Untitled72Untitled73Untitled74Untitled75Untitled76Untitled77Untitled78Untitled79Untitled80Untitled81Untitled82Untitled83Untitled84Untitled85Untitled86Untitled87Untitled88Untitled89Untitled90Untitled91Untitled92Untitled93Untitled94Untitled95Untitled96Untitled97Untitled98Untitled99Untitled100

%pyspark  
class2



FINISHED ▶ ⌵ 📖 ⚙️



default ▼

Took 0 sec. Last updated by anonymous at January 20 2017, 12:12:07 PM.

```
%pyspark
from io import StringIO
def show(p):
    img = StringIO()
    p.savefig(img, format='svg')
    img.seek(0)
    print( "%html " + img.read())
routes = sqlContext.sql("SELECT Route, count(*) as Count FROM air_traffic GROUP BY Route").co
route_freq = [(x[0],x[1]) for x in routes]
wordcloud = WordCloud().generate_from_frequencies(route_freq)
image = wordcloud.to_image()
image.show()
```

FINISHED ▶ ⌵ 📖 ⚙️

Took 2 min 24 sec. Last updated by anonymous at January 20 2017, 12:04:35 PM.

```
%pyspark
from io import StringIO
def show(p):
    img =StringIO()
    p.savefig(img, format='svg')
    img.seek(0)
    print( "%html " + img.read())
```

FINISHED ▶ ⌵ 📖 ⚙️

Took 0 sec. Last updated by anonymous at January 20 2017, 12:14:30 PM.

```
%pyspark
import matplotlib.pyplot as plt

#define some data
x = [1,2,3,4]
y = [20, 21, 20.5, 20.8]
```

FINISHED ▶ ⌵ 📖 ⚙️


Took 0 sec. Last updated by anonymous at January 20 2017, 12:14:32 PM.

```
%pyspark
plt.plot(x, y, linestyle="dashed", marker="o", color="green")
show(plt)
```

FINISHED ▶ ⌵ 📖 ⚙️



Avg Arrival Delay												
ABE	9.70	7.74	7.60	5.87	7.00	12.95	11.98	9.34	3.13	3.38	4.81	11.19
ABI	7.11	7.28	6.39	4.82	5.68	14.84	11.49	10.33	2.86	4.80	6.85	11.83
ABQ	6.84	7.13	6.75	5.28	5.52	9.62	8.21	7.24	2.66	5.37	4.45	10.74
ABY	8.73	11.08	6.15	1.60	4.22	12.88	14.79	14.37	6.63	8.29	6.79	11.58
ACK					3.56	26.33	24.23	19.51	2.89			
ACT	3.65	3.60	2.07	0.39	0.67	6.58	4.22	5.74	0.61	1.49	2.88	2.43
ACV	13.04	10.16	4.58	2.46	6.08	8.77	6.40	7.67	4.62	9.13	7.81	18.28
ACY	3.78	6.03	2.58	3.40	0.45	14.61	18.11	13.02	4.03	4.41	4.21	5.87
ADK	23.03	10.35	16.67	13.34	4.10	-0.30	-2.83	7.89	12.53	17.65	9.49	4.59
ADQ	11.33	8.33	7.97	4.14	5.36	9.47	10.02	8.92	7.18	6.29	7.00	13.50
AEX	5.38	8.01	5.69	4.13	3.95	14.44	12.21	10.85	3.05	3.89	5.80	9.76
AGS	7.03	9.39	8.91	8.43	7.71	15.18	17.69	13.38	6.43	6.13	6.71	9.33
AKN	12.08	16.00	16.82	16.90	8.45	14.93	8.55	10.69	10.21	13.36	6.17	14.37
ALB	9.72	8.77	9.34	7.01	8.16	13.95	14.19	12.24	3.89	4.61	5.24	11.27
ALO	-0.52	4.08	1.38	2.96	-4.20	7.27	5.55	5.94	-3.33	0.65	6.28	15.09
AMA	6.71	8.21	8.63	6.39	8.60	13.72	9.83	8.63	4.54	6.65	6.35	11.61
ANC	11.92	9.06	10.08	4.88	8.11	11.41	11.90	12.86	7.88	7.64	9.79	17.24
ANI	19.00	13.06	9.90	19.64	10.52	14.17	30.29	9.33	1.00	10.00	13.46	8.58
APF	9.60	10.02	12.19	5.77	4.12	15.13	27.15	22.19	4.67	3.13	9.16	8.81
ASE	18.70	16.16	10.50	1.69	-0.06	5.41	6.18	6.93	1.93	0.19	-3.16	23.05
ATL	8.44	9.97	8.65	5.18	5.40	11.75	13.32	10.03	4.41	5.62	6.72	12.63
ATW	14.09	16.72	10.83	7.87	4.33	12.32	10.57	9.39	3.33	3.52	6.33	20.55
AUS	6.63	6.61	7.03	5.03	6.01	10.38	9.04	7.36	2.08	4.43	4.71	9.68
AVL	5.64	5.62	4.95	4.23	4.11	12.32	13.23	12.01	4.24	6.21	4.24	9.57
AVP	13.64	10.37	8.93	7.91	8.86	16.60	14.68	12.03	4.61	3.93	5.26	15.07
AZO	13.04	10.85	8.52	5.83	5.62	9.87	7.97	6.13	1.93	2.15	5.38	15.28
BDL	7.44	7.16	7.91	5.65	6.35	12.58	11.54	9.95	2.84	3.19	3.74	8.69
BET	10.43	7.79	7.84	5.67	6.32	8.76	10.58	14.15	6.35	8.32	11.86	13.64
BFF												
BFL	7.39	5.20	3.80	1.29	2.16	4.41	3.56	2.47	1.76	3.92	2.83	10.91
3GM	10.13	8.25	4.78	4.77	5.75	11.42	7.87	7.11	2.25	2.10	4.06	11.62



30

http://localhost:8080/#/notebook/2C9P2XC6K

9/12

class2

Zeppelin

</

class Zeppelin


# class2

[illegible]

-30


Took 5 sec. Last updated by anonymous at January 20 2017, 12:14:39 PM.


READY ▶ 🔍 📖 ⚙️





# Zeppelin


class2




























default ▾

http://localhost:8080/#/notebook/2C9P2XC6K

12/12