# Programación para la Computación Científica - IA

# Data processing Cycle II

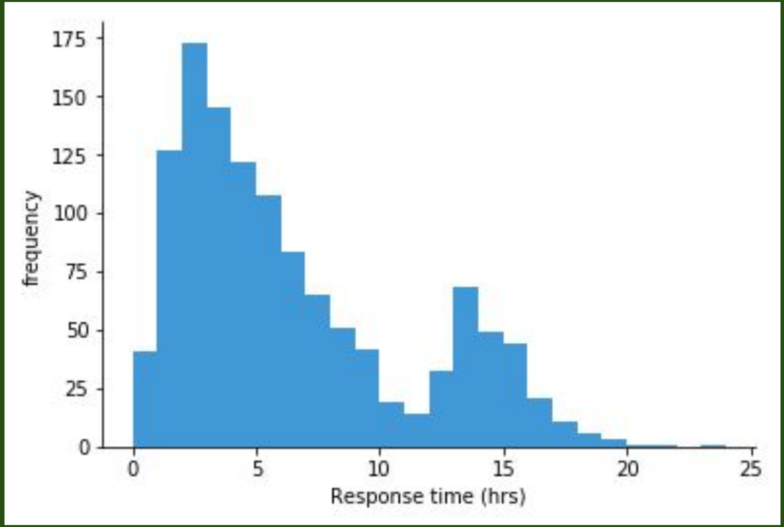
Universidad Sergio Arboleda

*Prof. John Corredor*

# Today goal’s

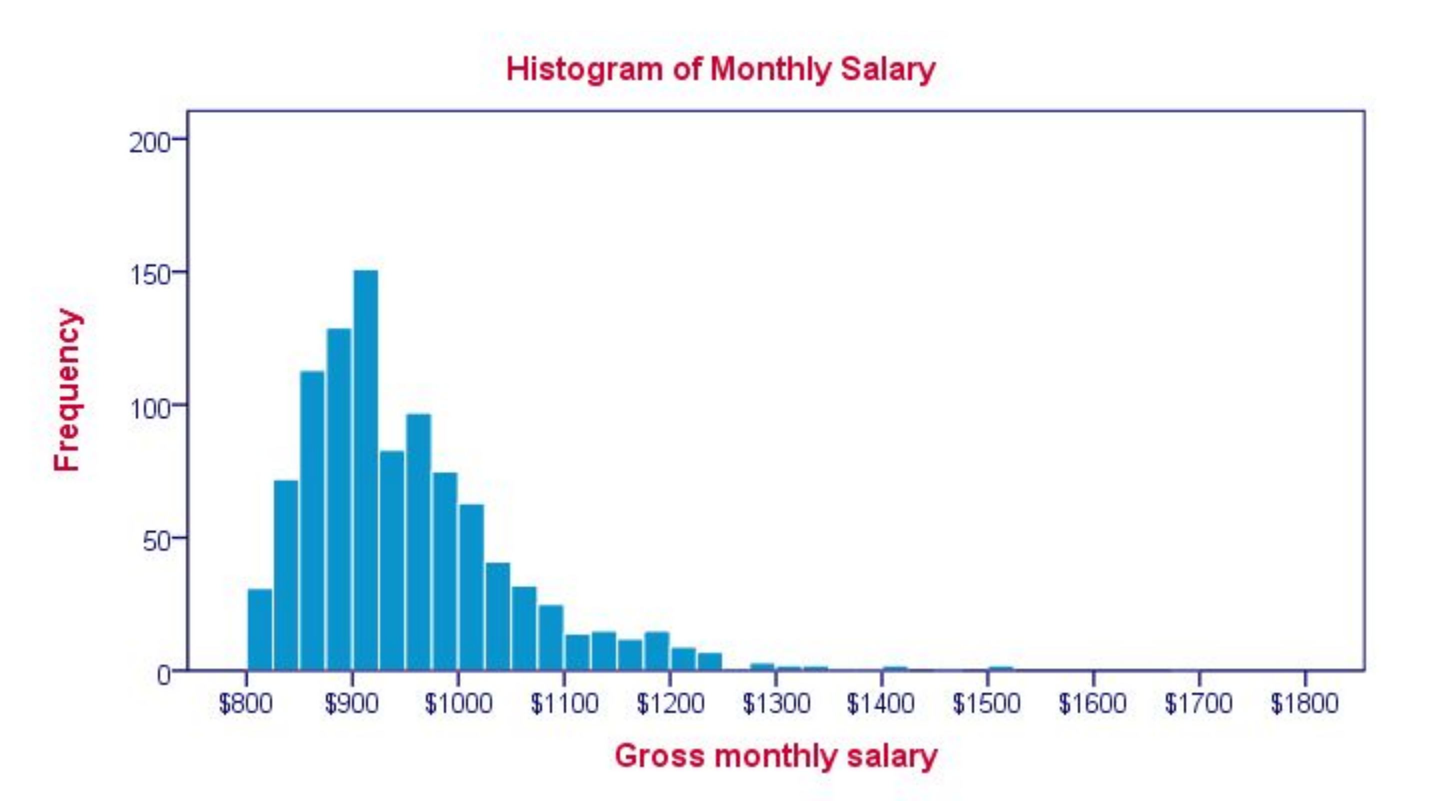
- Data processing - Histograms
- Introduction to Jupyter Notebook
- Principles of Pyhon

# What is a histogram?

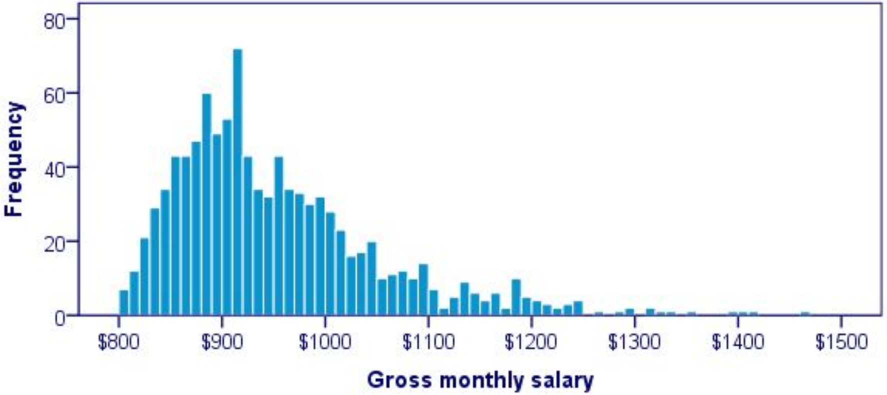


A histogram is a chart that plots the distribution of a numeric variable's values as a series of bars. Each bar typically covers a range of numeric values called a bin or class; a bar's height indicates the frequency of data points with a value within the corresponding bin.

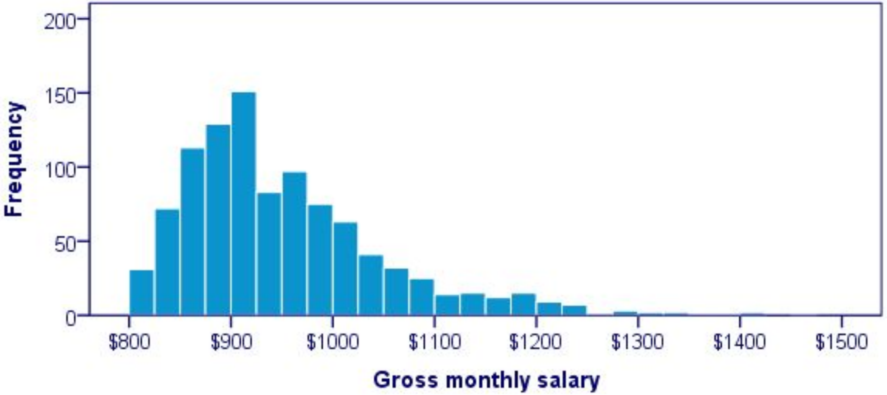
# What is a histogram?



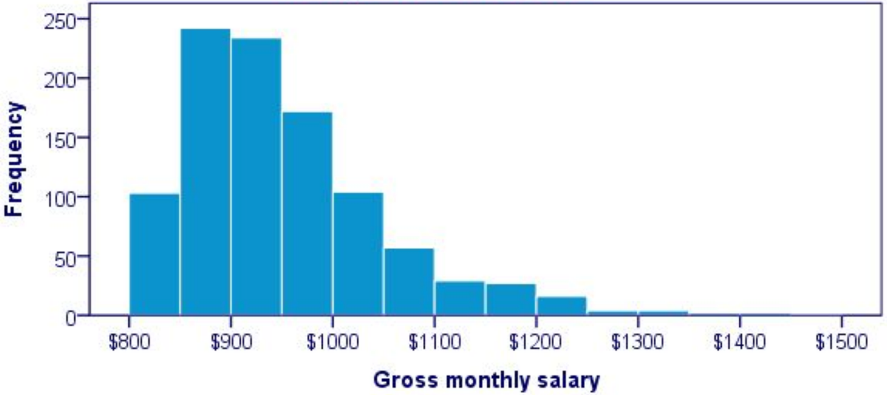
Histogram Bin Width = $10,-



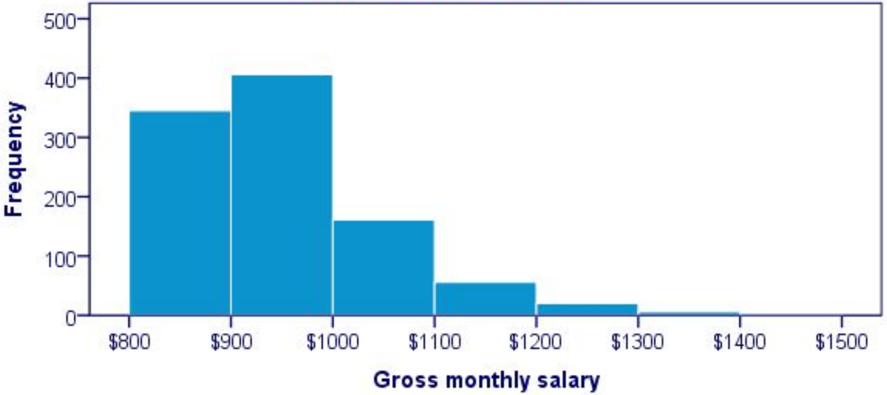
Histogram Bin Width = $25,-

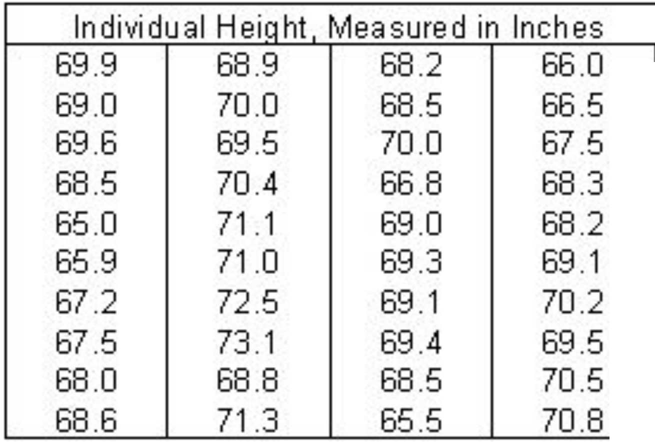


Histogram Bin Width = $50,-

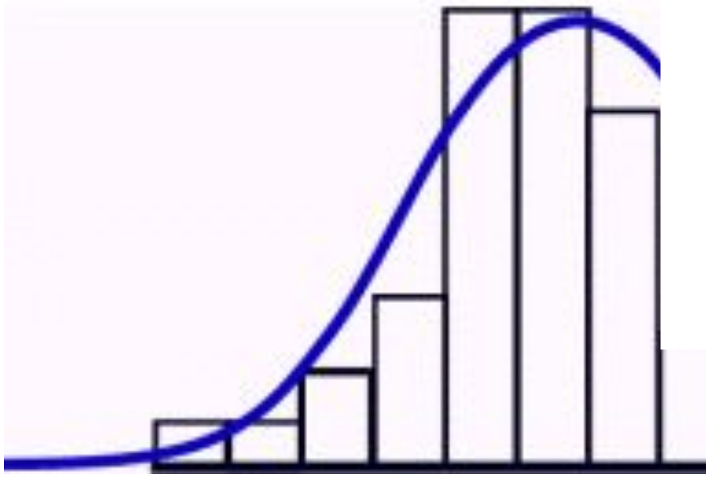


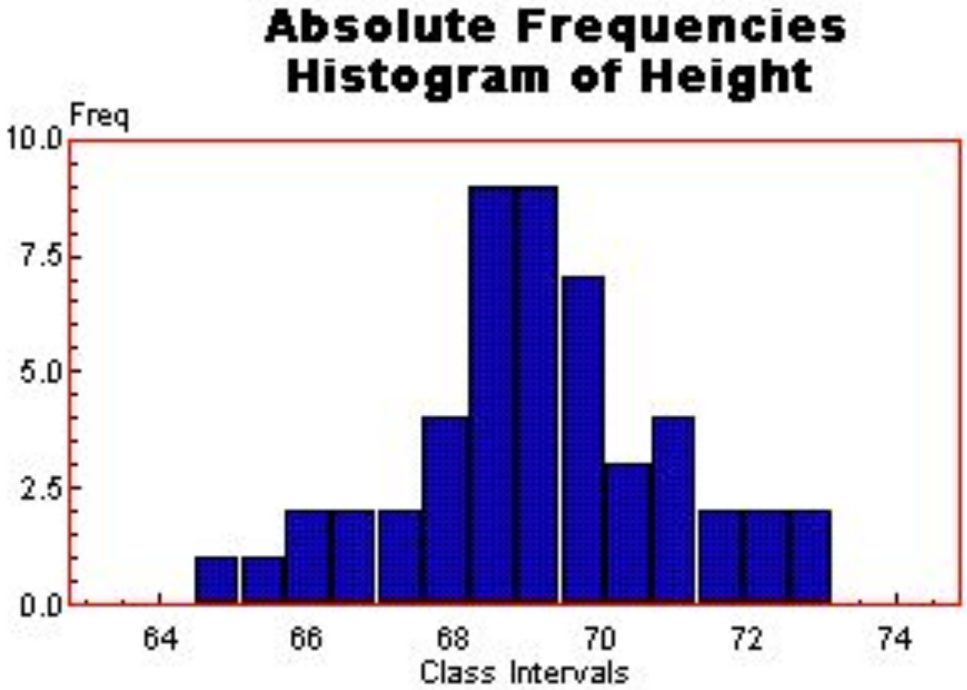
Histogram Bin Width = $100,-





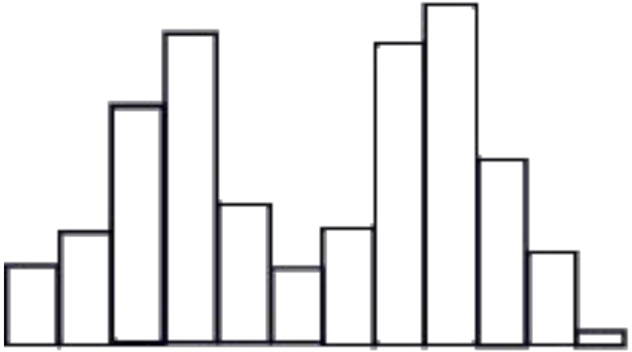
| Individual Height, Measured in Inches | | | | |
|---|---|---|---|---|
| 69.9 | 68.9 | 68.2 | 66.0 | 71.0 |
| 69.0 | 70.0 | 68.5 | 66.5 | |
| 69.6 | 69.5 | 70.0 | 67.5 | |
| 68.5 | 70.4 | 66.8 | 68.3 | |
| 65.0 | 71.1 | 69.0 | 68.2 | |
| 65.9 | 71.0 | 69.3 | 69.1 | |
| 67.2 | 72.5 | 69.1 | 70.2 | |
| 67.5 | 73.1 | 69.4 | 69.5 | |
| 68.0 | 68.8 | 68.5 | 70.5 | |
| 68.6 | 71.3 | 65.5 | 70.8 | |



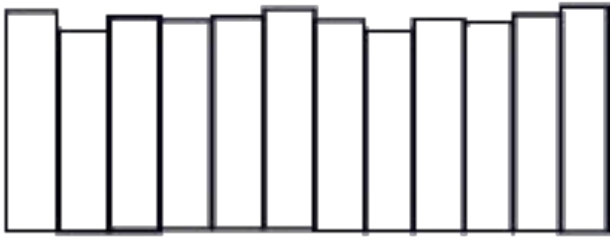


Absolute Frequencies Histogram of Height

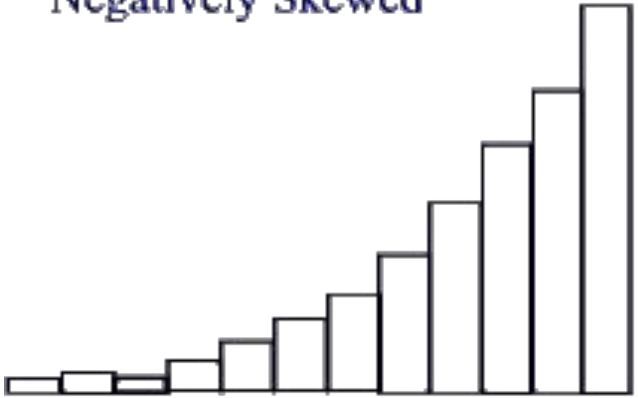
Bi-Modal Distribution



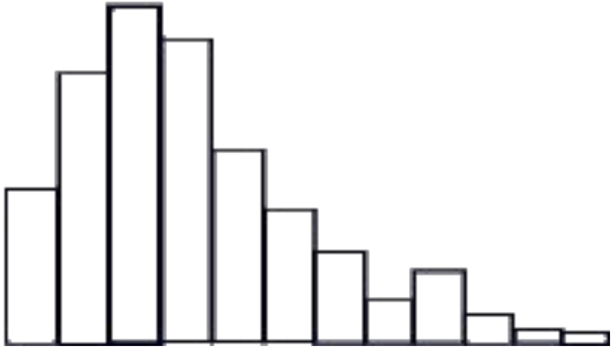
Unitary Distribution



Negatively Skewed



- Positively Skewed



# How to start?

1. Count the number of data points (50 in our height example) .
2. Determine the range of the sample - the difference between the highest and lowest values (73.1-65, or 8.1 inches in our height example.
3. Determine the number of class intervals. (Measurement System Analysis (MSA))

   You can use either of two methods as general guidelines in determining the number of intervals:

   A. Use ten intervals as a rule of thumb.

   B. Calculate the square root of the number of data points and round to the nearest whole number. In the case of our height example, the square root of 50 is 7.07, or 7 when rounded.

   You may wish to experiment with different interval numbers. If there are too many, the distribution will spread out, and the histogram will look flat. Likewise, if there are too few intervals, the distribution can look artificially tight.

4. Determine the interval class width by one of two methods:

0 = 0.81

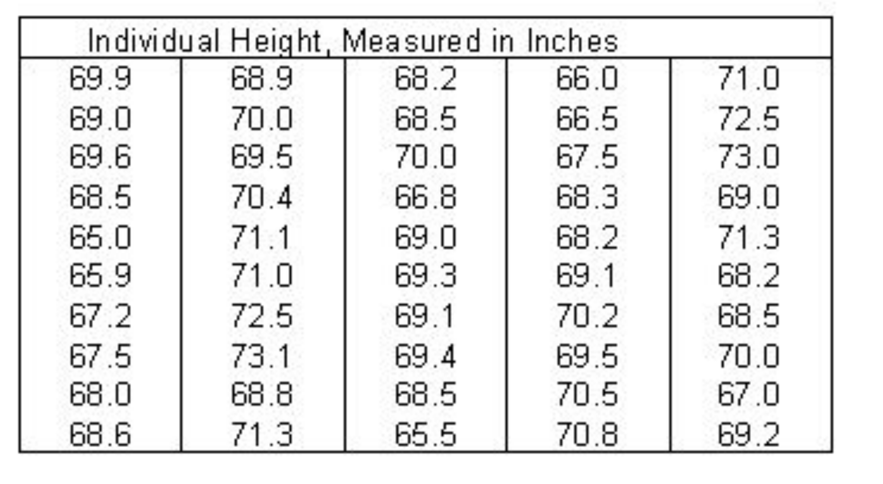
ree. In this case, the height data has a Standard

s

8

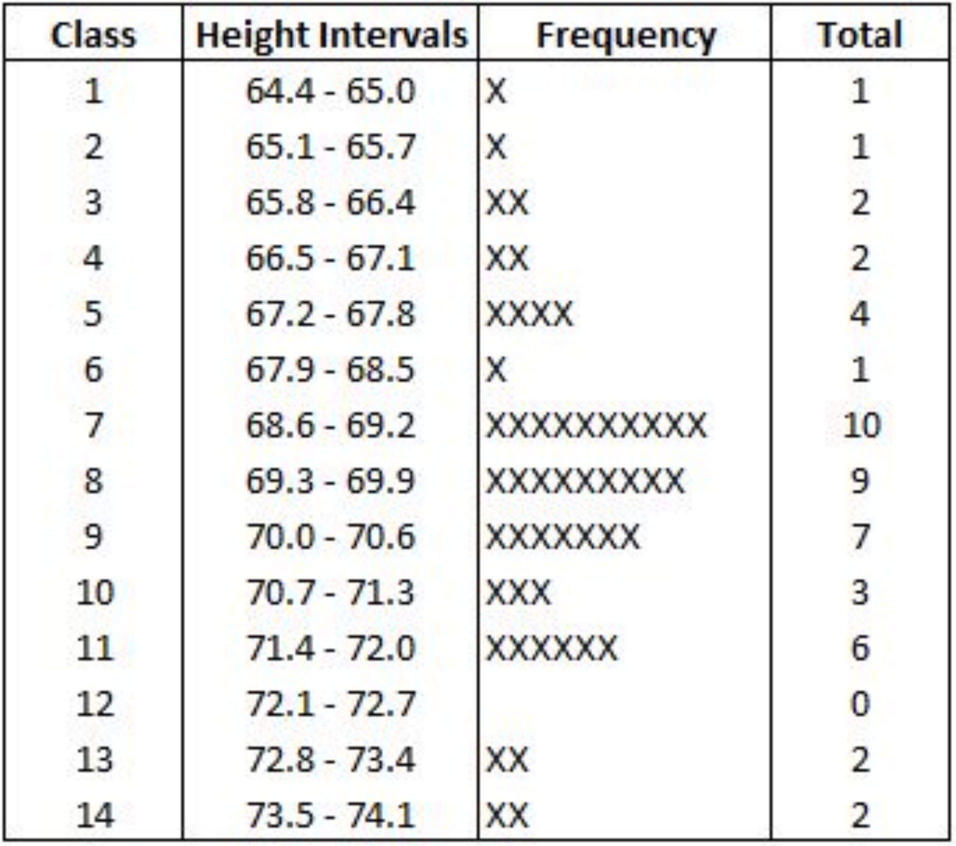
t

ti

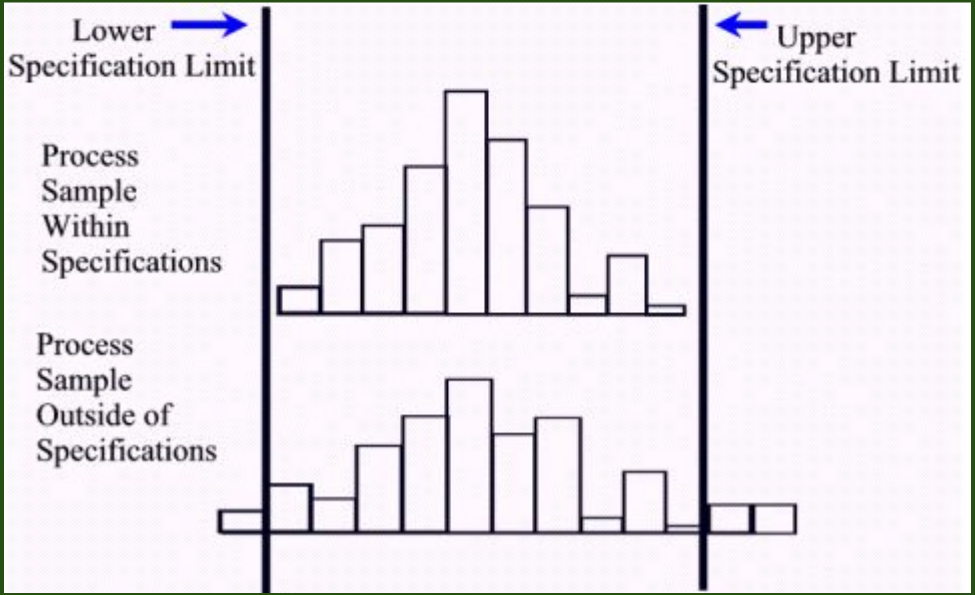


| Individual Height, Measured in Inches | | | | |
|---|---|---|---|---|
| 69.9 | 68.9 | 68.2 | 66.0 | 71.0 |
| 69.0 | 70.0 | 68.5 | 66.5 | 72.5 |
| 69.6 | 69.5 | 70.0 | 67.5 | 73.0 |
| 68.5 | 70.4 | 66.8 | 68.3 | 69.0 |
| 65.0 | 71.1 | 69.0 | 68.2 | 71.3 |
| 65.9 | 71.0 | 69.3 | 69.1 | 68.2 |
| 67.2 | 72.5 | 69.1 | 70.2 | 68.5 |
| 67.5 | 73.1 | 69.4 | 69.5 | 70.0 |
| 68.0 | 68.8 | 68.5 | 70.5 | 67.0 |
| 68.6 | 71.3 | 65.5 | 70.8 | 69.2 |

Once the histogram is developed, you can analyze the data with regard to customer expectations (specifications).

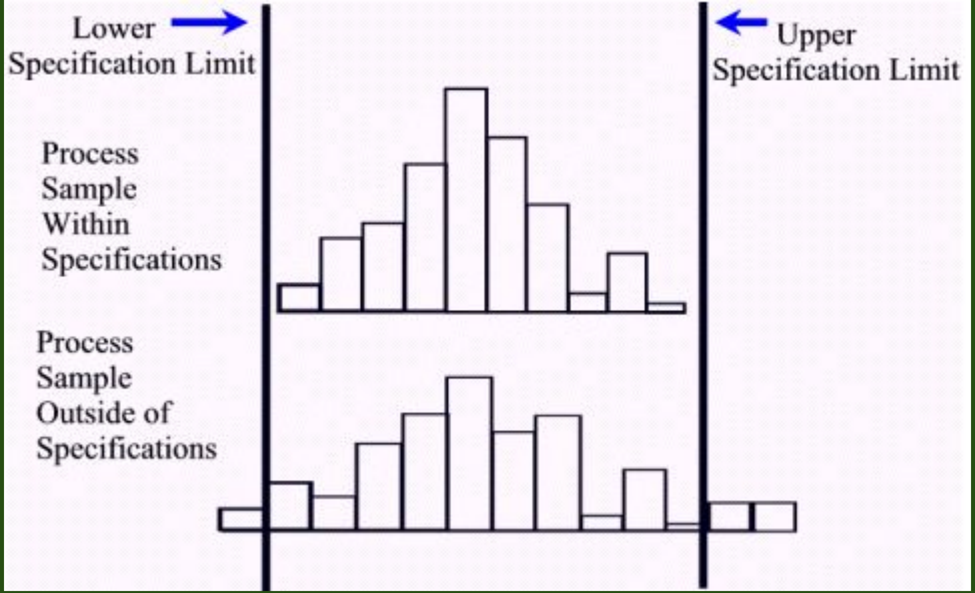


| Class | Height Intervals | Frequency | Total |
|---|---|---|---|
| 1 | 64.4 - 65.0 | X | 1 |
| 2 | 65.1 - 65.7 | X | 1 |
| 3 | 65.8 - 66.4 | XX | 2 |
| 4 | 66.5 - 67.1 | XX | 2 |
| 5 | 67.2 - 67.8 | XXXX | 4 |
| 6 | 67.9 - 68.5 | X | 1 |
| 7 | 68.6 - 69.2 | XXXXXXXXXX | 10 |
| 8 | 69.3 - 69.9 | XXXXXXXXX | 9 |
| 9 | 70.0 - 70.6 | XXXXXXX | 7 |
| 10 | 70.7 - 71.3 | XXX | 3 |
| 11 | 71.4 - 72.0 | XXXXXX | 6 |
| 12 | 72.1 - 72.7 | | 0 |
| 13 | 72.8 - 73.4 | XX | 2 |
| 14 | 73.5 - 74.1 | XX | 2 |



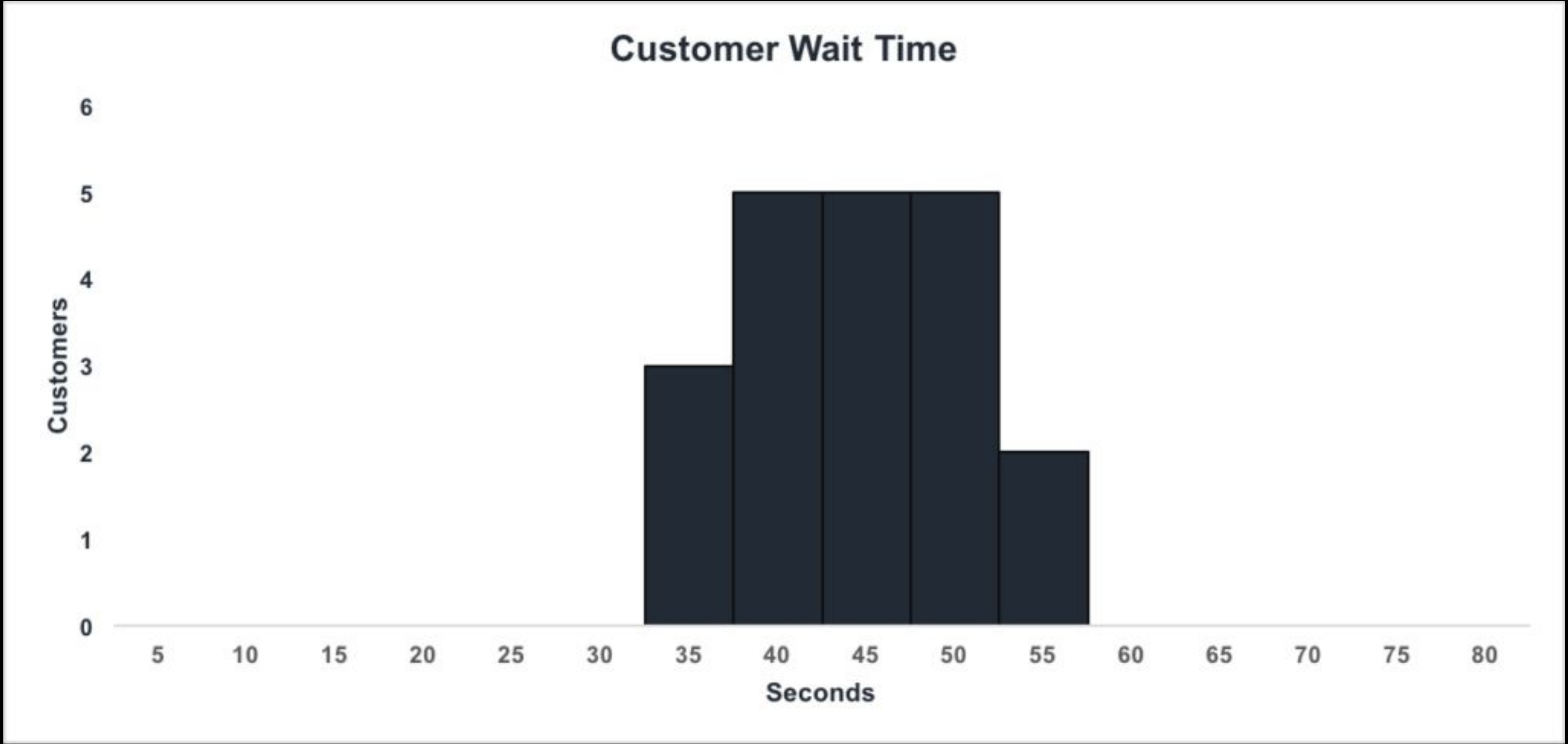
The first histogram of a process sample falls within the specifications, while the second has a portion of the histogram outside of the specifications.

The second histogram has too much dispersion, or variability, to meet customer expectations. The indication is that action must be taken to make the output more consistent, or some number of defects will be produced. (Statistical Process Control)

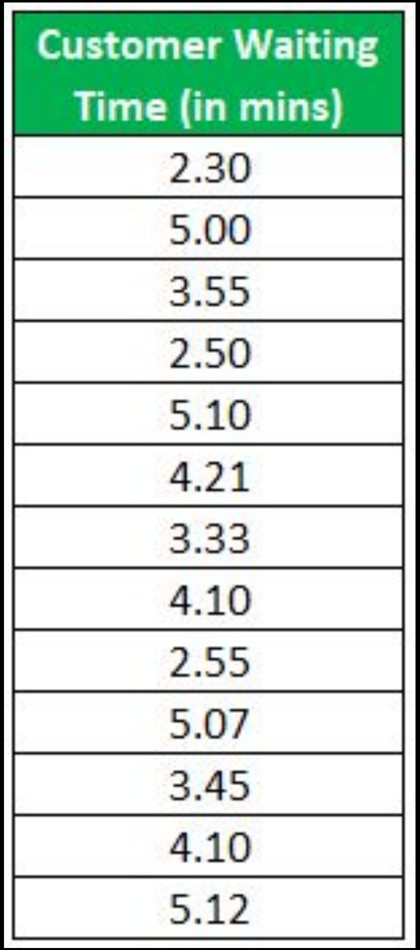


The first histogram of a process sample falls within the specifications, while the second has a portion of the histogram outside of the specifications.

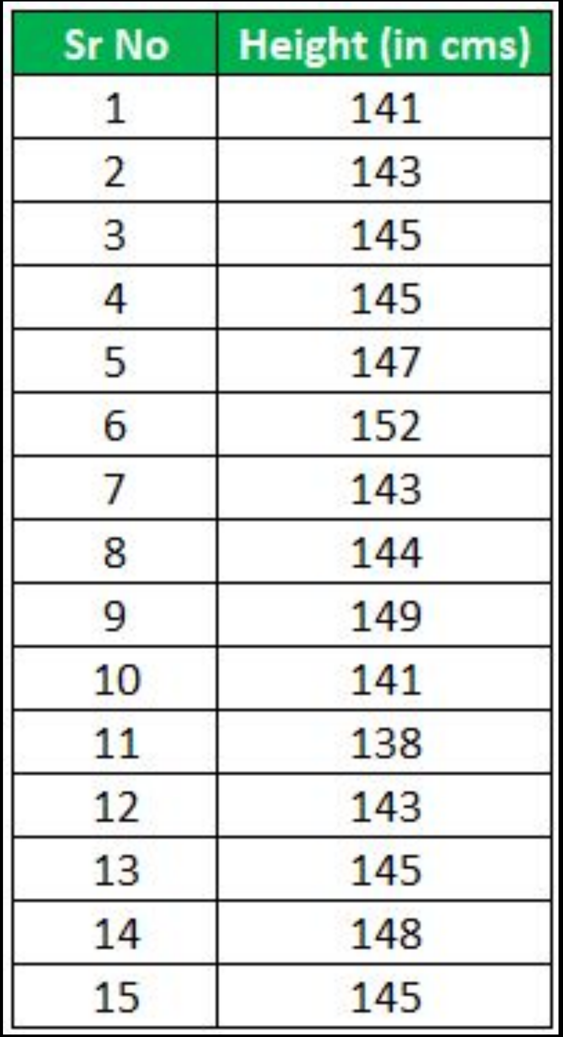
The second histogram has too much dispersion, or variability, to meet customer expectations. The indication is that action must be taken to make the output more consistent, or some number of defects will be produced. (Statistical Process Control)


Customer Wait Time
Customers
6
5
4
3
2
1
0
5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
Seconds

# Challenge 01



| Customer Waiting Time (in mins) |
|---|
| 2.30 |
| 5.00 |
| 3.55 |
| 2.50 |
| 5.10 |
| 4.21 |
| 3.33 |
| 4.10 |
| 2.55 |
| 5.07 |
| 3.45 |
| 4.10 |
| 5.12 |

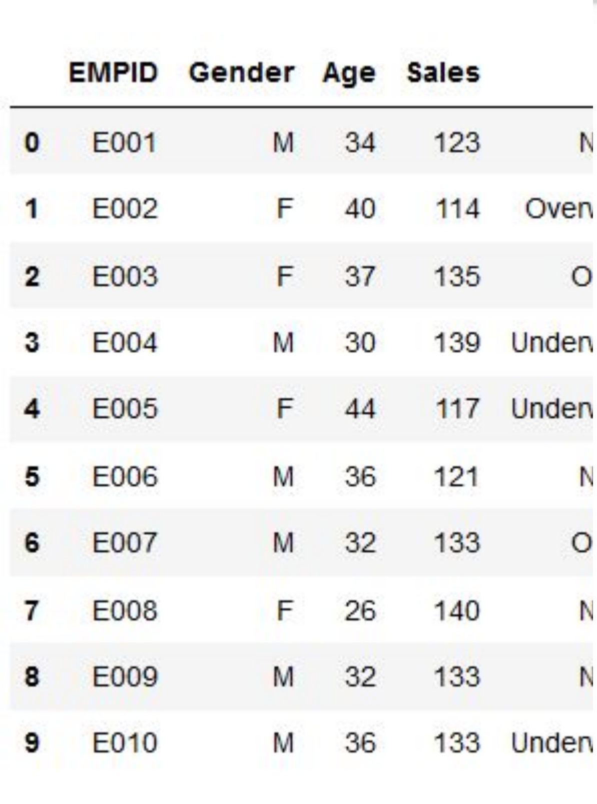


| Sr No | Height (in cms) |
|---|---|
| 1 | 141 |
| 2 | 143 |
| 3 | 145 |
| 4 | 145 |
| 5 | 147 |
| 6 | 152 |
| 7 | 143 |
| 8 | 144 |
| 9 | 149 |
| 10 | 141 |
| 11 | 138 |
| 12 | 143 |
| 13 | 145 |
| 14 | 148 |
| 15 | 145 |

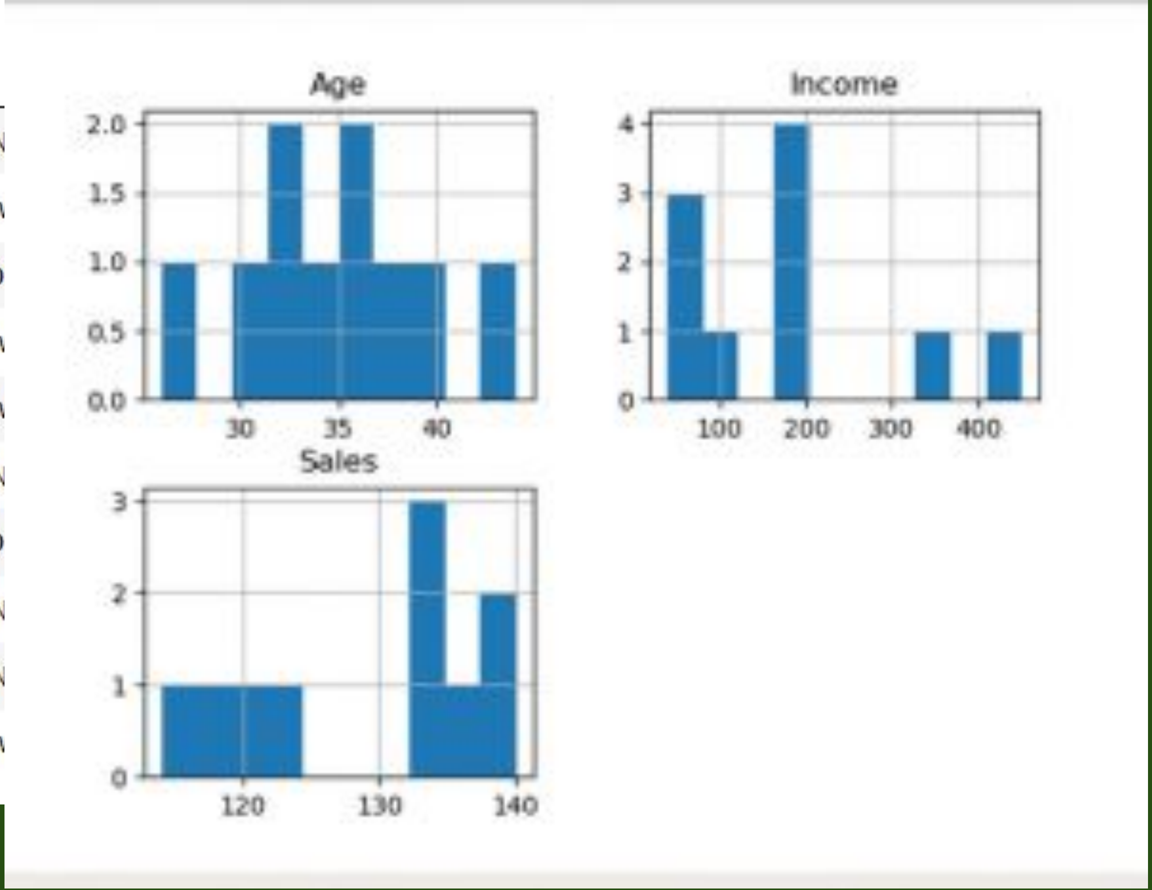
# Challenge 02

El Sr. Larry, un médico famoso, está realizando una investigación sobre la altura de los estudiantes que estudian en el octavo estándar. Ha reunido una muestra de 15 estudiantes, pero quiere saber cuál es la categoría máxima a la que pertenecen.

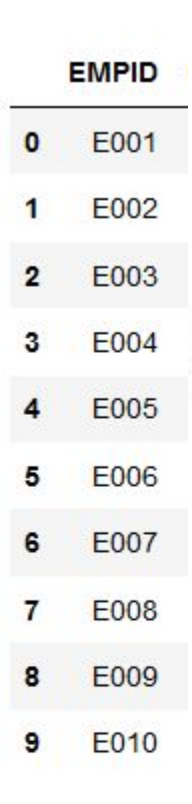
# Data visualization with different Charts



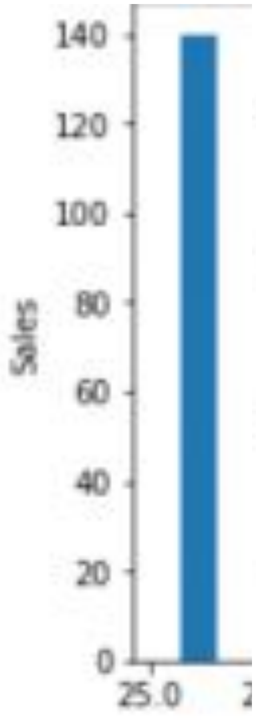
|   | EMPID | Gender | Age | Sales |   |
|---|---|---|---|---|---|
| **0** | E001 | M | 34 | 123 | N |
| **1** | E002 | F | 40 | 114 | Over |
| **2** | E003 | F | 37 | 135 | O |
| **3** | E004 | M | 30 | 139 | Under |
| **4** | E005 | F | 44 | 117 | Under |
| **5** | E006 | M | 36 | 121 | N |
| **6** | E007 | M | 32 | 133 | O |
| **7** | E008 | F | 26 | 140 | N |
| **8** | E009 | M | 32 | 133 | N |
| **9** | E010 | M | 36 | 133 | Under |

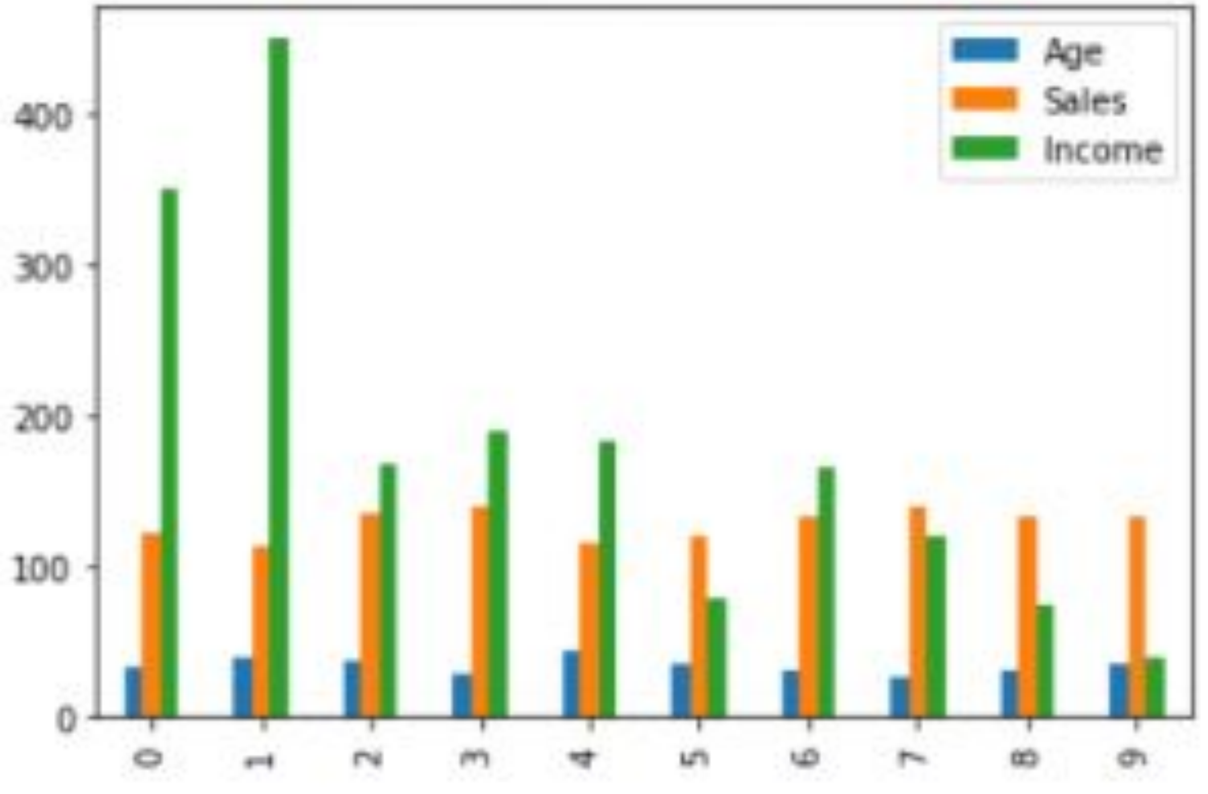


# Data visualization with different Charts

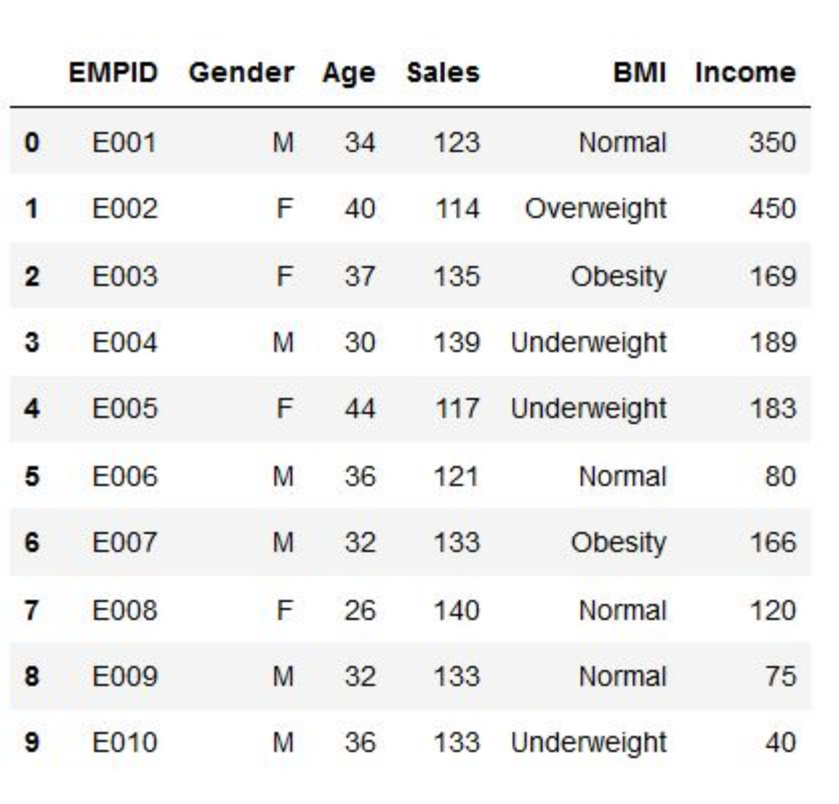


|  | EMPID |
|---|---|
| **0** | E001 |
| **1** | E002 |
| **2** | E003 |
| **3** | E004 |
| **4** | E005 |
| **5** | E006 |
| **6** | E007 |
| **7** | E008 |
| **8** | E009 |
| **9** | E010 |

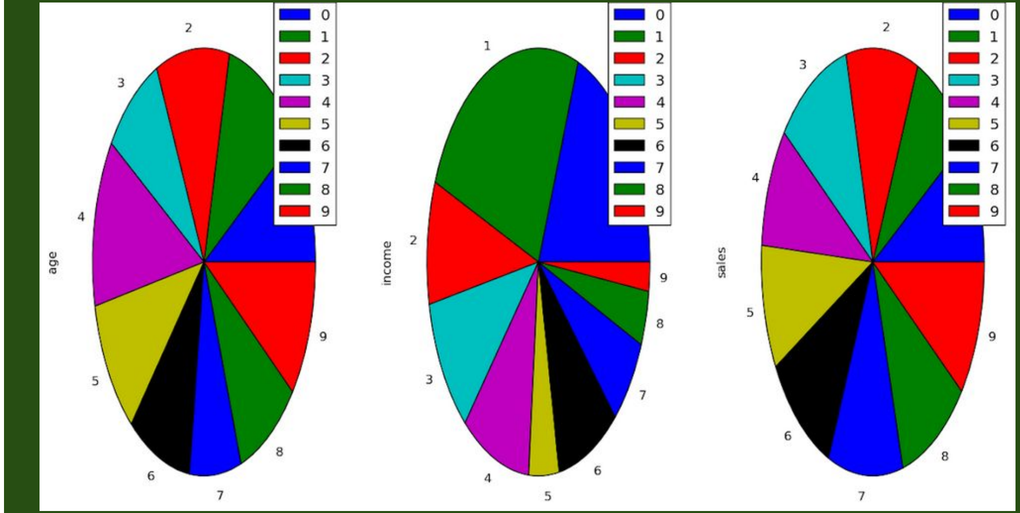




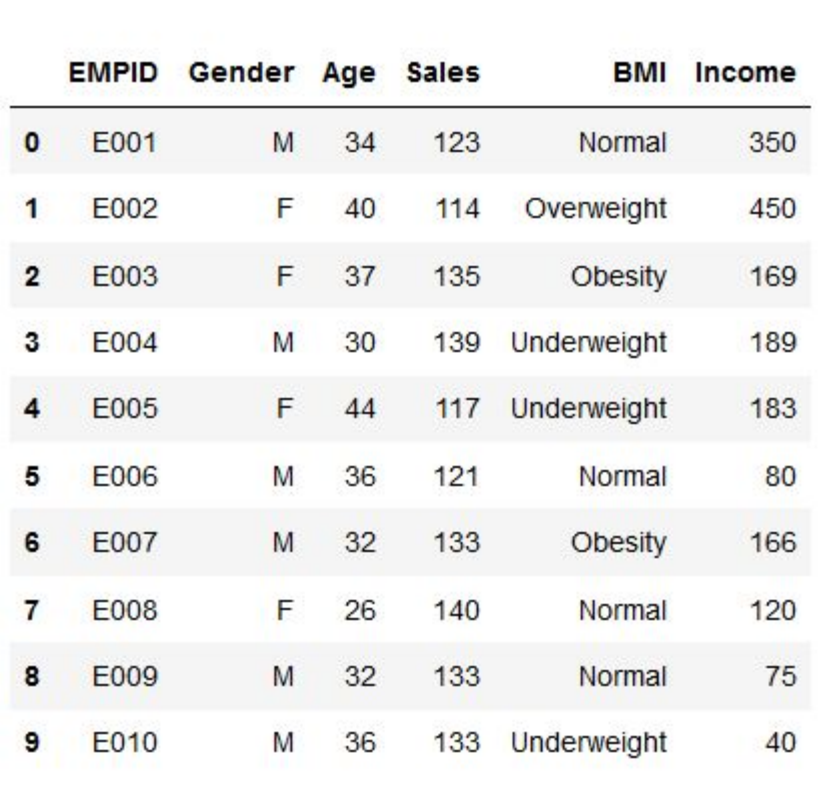
# Data visualization with different Charts



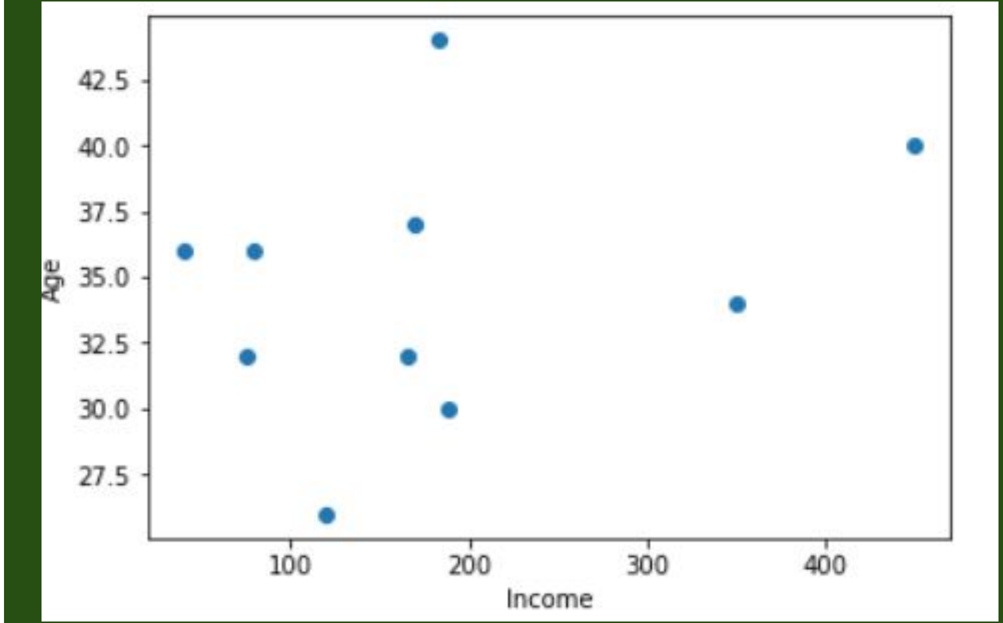
| | EMPID | Gender | Age | Sales | BMI | Income |
|---|---|---|---|---|---|---|
| **0** | E001 | M | 34 | 123 | Normal | 350 |
| **1** | E002 | F | 40 | 114 | Overweight | 450 |
| **2** | E003 | F | 37 | 135 | Obesity | 169 |
| **3** | E004 | M | 30 | 139 | Underweight | 189 |
| **4** | E005 | F | 44 | 117 | Underweight | 183 |
| **5** | E006 | M | 36 | 121 | Normal | 80 |
| **6** | E007 | M | 32 | 133 | Obesity | 166 |
| **7** | E008 | F | 26 | 140 | Normal | 120 |
| **8** | E009 | M | 32 | 133 | Normal | 75 |
| **9** | E010 | M | 36 | 133 | Underweight | 40 |



# Data visualization with different Charts



| | EMPID | Gender | Age | Sales | BMI | Income |
|---|---|---|---|---|---|---|
| **0** | E001 | M | 34 | 123 | Normal | 350 |
| **1** | E002 | F | 40 | 114 | Overweight | 450 |
| **2** | E003 | F | 37 | 135 | Obesity | 169 |
| **3** | E004 | M | 30 | 139 | Underweight | 189 |
| **4** | E005 | F | 44 | 117 | Underweight | 183 |
| **5** | E006 | M | 36 | 121 | Normal | 80 |
| **6** | E007 | M | 32 | 133 | Obesity | 166 |
| **7** | E008 | F | 26 | 140 | Normal | 120 |
| **8** | E009 | M | 32 | 133 | Normal | 75 |
| **9** | E010 | M | 36 | 133 | Underweight | 40 |



# Why should I prepare my data?

- Garbage in, garbage out
- Reduce errors
- Remove duplicate records
- Fix missing values
- Correct range values
- Fix formatting (i.e. date, text, number)

# Experience Check

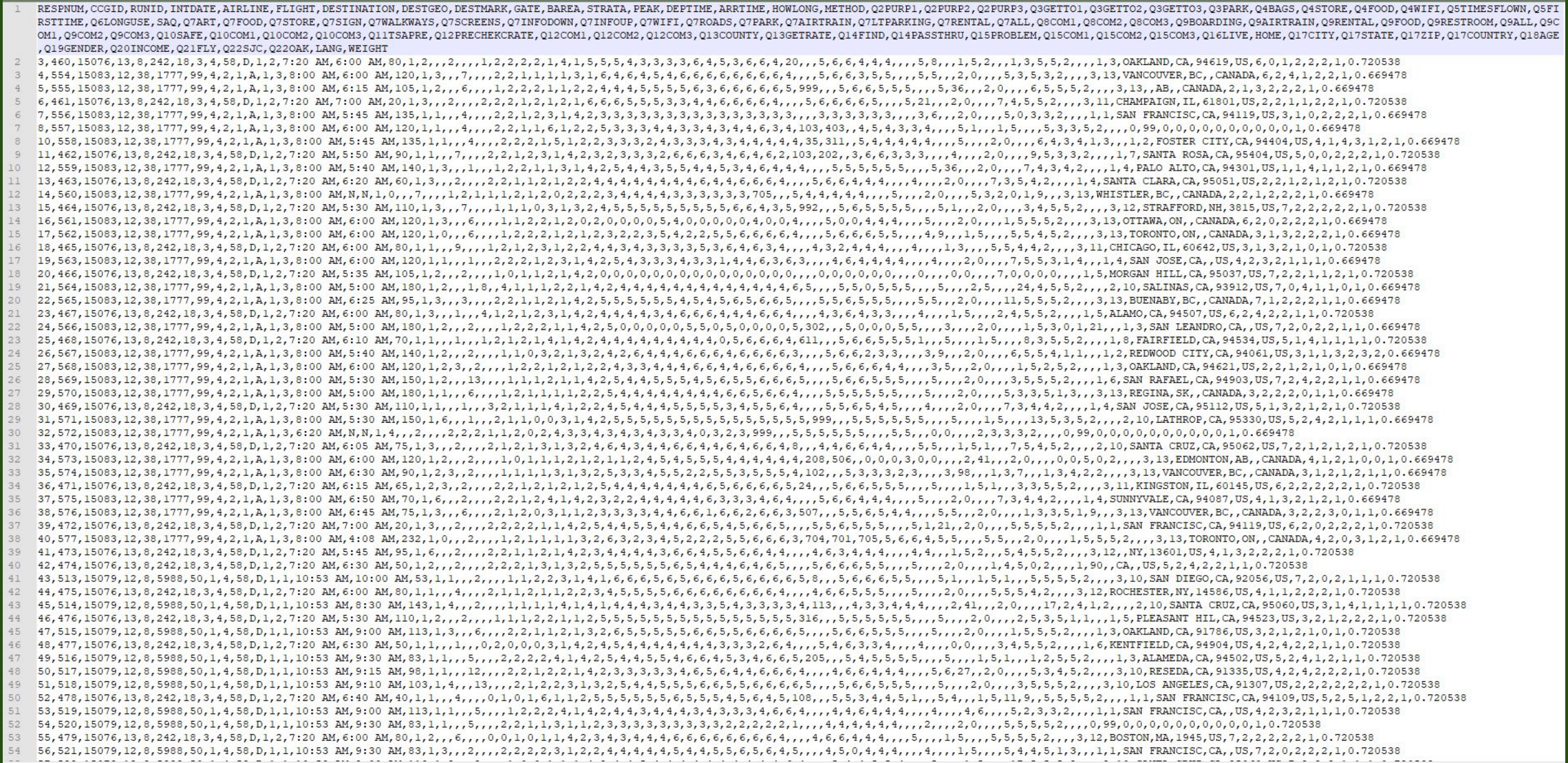
- How many people have experience with Python?
- What types of data formats do you use in your organizations?
    - CSV, Excel, PDF, JSON, XML, SQL databases, etc
- What types of tools do you use?
    - Hoja de Calculo, ACL, IDEA, SQL Server, Python, R, SAS, Cognos, etc

# What types of data formats might I encounter?

- Comma Separated Value (CSV)
- Hoja de calculo
- JavaScript Object Notation (JSON)
- Structured Query Language (SQL)
- And more!

# CVS example

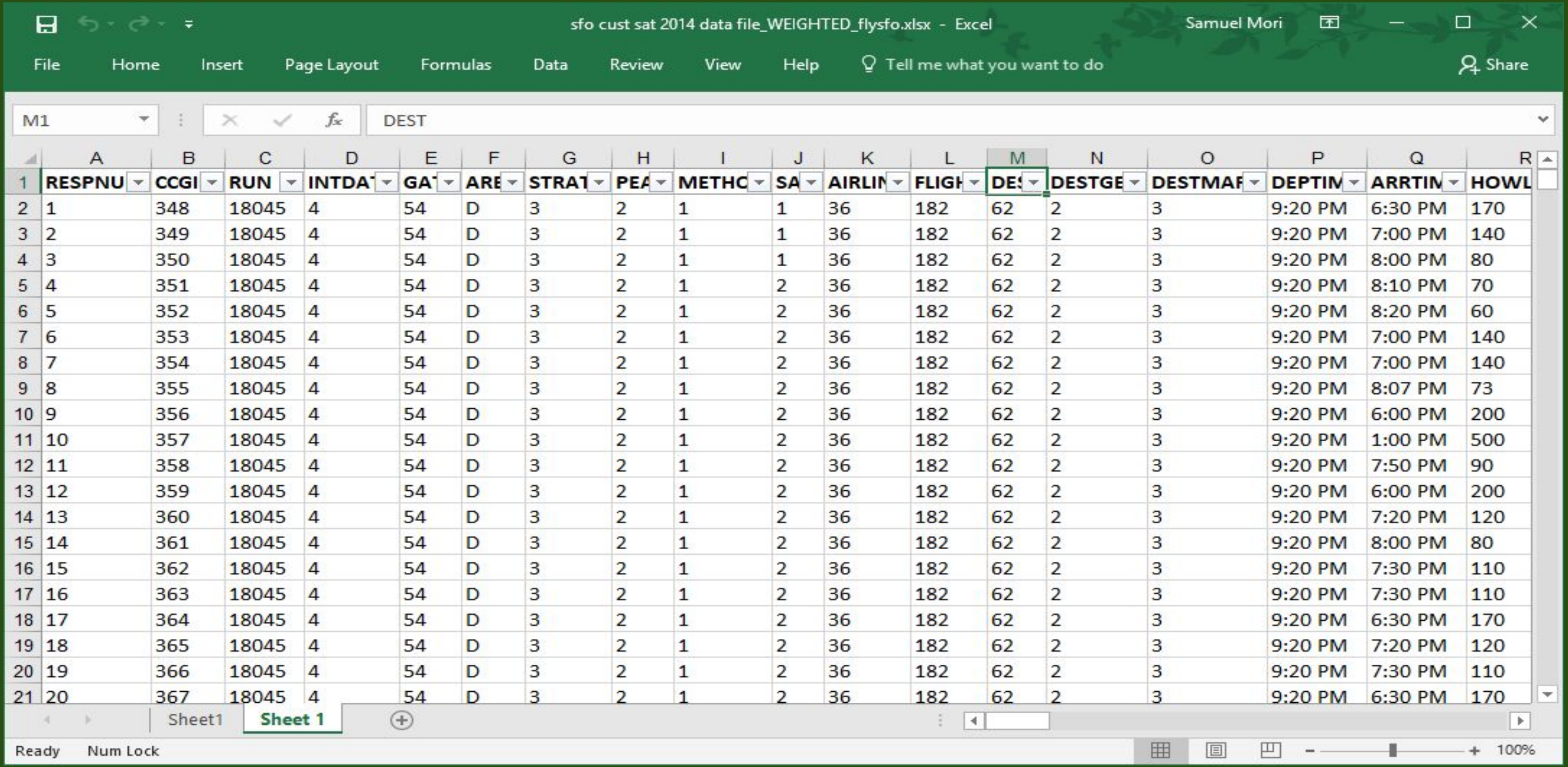
- SFO Airport Survey Results



```
RESPNUM,CCGID,RUNID,INTDATE,AIRLINE,FLIGHT,DESTINATION,DESTGEO,DESTMARK,GATE,BAREA,STRATA,PEAK,DEPTIME,ARRTIME,HOWLONG,METHOD,Q2PURP1,Q2PURP2,Q2PURP3,Q3GETTO1,Q3GETTO2,Q3GETTO3,Q3PARK,Q4BAGS,Q4STORE,Q4FOOD,Q4WIFI,Q5TIMESFLOWN,Q5FI
RSTTIME,Q6LONGUSE,SAQ,Q7ART,Q7FOOD,Q7STORE,Q7SIGN,Q7WALKWAYS,Q7SCREENS,Q7INFODOWN,Q7INFOUP,Q7WIFI,Q7ROADS,Q7PARK,Q7AIRTRAIN,Q7LTPARKING,Q7RENTAL,Q7ALL,Q8COM1,Q8COM2,Q8COM3,Q9BOARDING,Q9AIRTRAIN,Q9RENTAL,Q9FOOD,Q9RESTROOM,Q9ALL,Q9C
OM1,Q9COM2,Q9COM3,Q10SAFE,Q10COM1,Q10COM2,Q10COM3,Q11TSAPRE,Q12PRECHEKCRATE,Q12COM1,Q12COM2,Q12COM3,Q13COUNTY,Q13GETRATE,Q14FIND,Q14PASSTHRU,Q15PROBLEM,Q15COM1,Q15COM2,Q15COM3,Q16LIVE,HOME,Q17CITY,Q17STATE,Q17ZIP,Q17COUNTRY,Q18AGE
,Q19GENDER,Q20INCOME,Q21FLY,Q22SJC,Q22OAK,LANG,WEIGHT
3,460,15076,13,8,242,18,3,4,58,D,1,2,7:20 AM,6:00 AM,80,1,2,,,2,,,,1,2,2,2,2,1,4,1,5,5,5,4,3,3,3,3,6,4,5,3,6,6,4,20,,,5,6,6,4,4,4,,,,5,8,,,1,5,2,,,1,3,5,5,2,,,,1,3,OAKLAND,CA,94619,US,6,0,1,2,2,2,1,0.720538
4,554,15083,12,38,1777,99,4,2,1,A,1,3,8:00 AM,6:00 AM,120,1,3,,,7,,,,2,2,1,1,1,1,3,1,6,4,6,4,5,4,6,6,6,6,6,6,6,6,4,,,,5,6,6,3,5,5,,,,5,5,,,2,0,,,,5,3,5,3,2,,,,3,13,VANCOUVER,BC,,CANADA,6,2,4,1,2,2,1,0.669478
5,555,15083,12,38,1777,99,4,2,1,A,1,3,8:00 AM,6:15 AM,105,1,2,,,6,,,,1,2,2,2,1,1,2,2,4,4,4,5,5,5,5,6,3,6,6,6,6,6,5,999,,,5,6,6,5,5,5,,,,5,36,,,2,0,,,,6,5,5,5,2,,,,3,13,,AB,,CANADA,2,1,3,2,2,2,1,0.669478
6,461,15076,13,8,242,18,3,4,58,D,1,2,7:20 AM,7:00 AM,20,1,3,,,2,,,,2,2,2,1,2,1,2,1,6,6,6,5,5,5,3,3,4,4,6,6,6,6,4,,,,5,6,6,6,6,5,,,,5,21,,,2,0,,,,7,4,5,5,2,,,,3,11,CHAMPAIGN,IL,61801,US,2,2,1,1,2,2,1,0.720538
7,556,15083,12,38,1777,99,4,2,1,A,1,3,8:00 AM,5:45 AM,135,1,1,,,4,,,,2,2,1,2,3,1,4,2,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,,,,3,3,3,3,3,3,,,,3,6,,,2,0,,,,5,0,3,3,2,,,,1,1,SAN FRANCISC,CA,94119,US,3,1,0,2,2,2,1,0.669478
8,557,15083,12,38,1777,99,4,2,1,A,1,3,8:00 AM,6:00 AM,120,1,1,,,4,,,,2,2,1,1,6,1,2,2,5,3,3,3,4,4,3,3,4,3,4,4,6,3,4,103,403,,4,5,4,3,3,4,,,,5,1,,,1,5,,,,5,3,3,5,2,,,,0,99,0,0,0,0,0,0,0,0,0,0,1,0.669478
10,558,15083,12,38,1777,99,4,2,1,A,1,3,8:00 AM,5:45 AM,135,1,1,,,4,,,,2,2,2,1,5,1,2,2,3,3,3,2,4,3,3,3,4,3,4,4,4,4,4,35,311,,5,4,4,4,4,4,,,,5,,,,2,0,,,,6,4,3,4,1,3,,,1,2,FOSTER CITY,CA,94404,US,4,1,4,3,1,2,1,0.669478
11,462,15076,13,8,242,18,3,4,58,D,1,2,7:20 AM,5:50 AM,90,1,1,,,7,,,,2,2,1,2,3,1,4,2,3,2,3,3,3,2,6,6,6,3,4,6,4,6,2,103,202,,3,6,6,3,3,3,,,,4,,,,2,0,,,,9,5,3,3,2,,,,1,7,SANTA ROSA,CA,95404,US,5,0,0,2,2,2,1,0.720538
12,559,15083,12,38,1777,99,4,2,1,A,1,3,8:00 AM,5:40 AM,140,1,3,,,1,,,,1,2,2,1,1,3,1,4,2,5,4,4,3,5,5,4,4,5,3,4,6,4,4,4,,,,5,5,5,5,5,5,,,,5,36,,,2,0,,,,7,4,3,4,2,,,,1,4,PALO ALTO,CA,94301,US,1,1,4,1,1,2,1,0.669478
13,463,15076,13,8,242,18,3,4,58,D,1,2,7:20 AM,6:20 AM,60,1,3,,,2,,,,2,2,1,1,2,1,2,2,4,4,4,4,4,4,4,4,6,4,4,6,6,6,4,,,,5,6,6,4,4,4,,,,4,,,,2,0,,,,7,3,5,4,2,,,,1,4,SANTA CLARA,CA,95051,US,2,2,1,2,1,2,1,0.720538
14,560,15083,12,38,1777,99,4,2,1,A,1,3,8:00 AM,N,N,1,0,,,7,,,,1,2,1,1,1,2,1,2,0,2,2,2,3,4,4,4,4,3,3,3,3,3,3,705,,,5,4,4,4,4,4,,,,5,,,,2,0,,,,5,3,2,0,1,9,,,3,13,WHISTLER,BC,,CANADA,2,2,1,2,2,2,1,0.669478
15,464,15076,13,8,242,18,3,4,58,D,1,2,7:20 AM,5:30 AM,110,1,3,,,7,,,,1,1,1,0,3,1,3,2,4,5,5,5,5,5,5,5,5,6,6,4,3,5,992,,,5,6,5,5,5,5,,,,5,1,,,2,0,,,,3,4,5,5,2,,,,3,12,STRAFFORD,NH,3815,US,7,2,2,2,2,2,1,0.720538
16,561,15083,12,38,1777,99,4,2,1,A,1,3,8:00 AM,6:00 AM,120,1,3,,,6,,,,1,1,2,2,1,2,0,2,0,0,0,0,5,4,0,0,0,0,0,4,0,0,4,,,,5,0,0,4,4,4,,,,5,,,,2,0,,,,1,5,5,5,2,,,,3,13,OTTAWA,ON,,CANADA,6,2,0,2,2,2,1,0.669478
17,562,15083,12,38,1777,99,4,2,1,A,1,3,8:00 AM,6:00 AM,120,1,0,,,6,,,,1,2,2,2,1,2,1,2,3,2,2,3,5,4,2,2,5,5,6,6,6,6,4,,,,5,6,6,6,5,5,,,,4,9,,,1,5,,,,5,5,4,5,2,,,,3,13,TORONTO,ON,,CANADA,3,1,3,2,2,2,1,0.669478
18,465,15076,13,8,242,18,3,4,58,D,1,2,7:20 AM,6:00 AM,80,1,1,,,9,,,,1,2,1,2,3,1,2,2,4,4,3,4,3,3,3,3,5,3,6,4,6,3,4,,,,4,3,2,4,4,4,,,,4,,,,1,3,,,,5,5,4,4,2,,,,3,11,CHICAGO,IL,60642,US,3,1,3,2,1,0,1,0.720538
19,563,15083,12,38,1777,99,4,2,1,A,1,3,8:00 AM,6:00 AM,120,1,1,,,1,,,2,2,2,1,2,3,1,4,2,5,4,3,3,3,4,3,3,1,4,4,6,3,6,3,,,,4,6,4,4,4,4,,,,4,,,,2,0,,,,7,5,5,3,1,4,,,1,4,SAN JOSE,CA,,US,4,2,3,2,1,1,1,0.669478
20,466,15076,13,8,242,18,3,4,58,D,1,2,7:20 AM,5:35 AM,105,1,2,,,2,,,,1,0,1,1,2,1,4,2,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,,,,0,0,0,0,0,0,,,,0,,,,0,0,,,,7,0,0,0,0,,,,1,5,MORGAN HILL,CA,95037,US,7,2,2,1,1,2,1,0.720538
21,564,15083,12,38,1777,99,4,2,1,A,1,3,8:00 AM,5:00 AM,180,1,2,,,1,8,,4,1,1,1,2,2,1,4,2,4,4,4,4,4,4,4,4,4,4,4,4,4,6,5,,,,5,5,0,5,5,5,,,,5,,,,2,5,,,,24,4,5,5,2,,,,2,10,SALINAS,CA,93912,US,7,0,4,1,1,0,1,0.669478
22,565,15083,12,38,1777,99,4,2,1,A,1,3,8:00 AM,6:25 AM,95,1,3,,,3,,,,2,2,1,1,2,1,4,2,5,5,5,5,5,5,4,5,4,5,6,5,6,6,5,,,,5,5,6,5,5,5,,,,5,5,,,2,0,,,,11,5,5,5,2,,,,3,13,BUENABY,BC,,CANADA,7,1,2,2,2,1,1,0.669478
23,467,15076,13,8,242,18,3,4,58,D,1,2,7:20 AM,6:00 AM,80,1,3,,,1,,,4,1,2,1,2,3,1,4,2,4,4,4,4,3,4,6,6,6,4,4,4,6,6,4,,,,4,3,6,4,3,3,,,,4,,,,1,5,,,,2,4,5,5,2,,,,1,5,ALAMO,CA,94507,US,6,2,4,2,2,1,1,0.720538
24,566,15083,12,38,1777,99,4,2,1,A,1,3,8:00 AM,5:00 AM,180,1,2,,,2,,,,1,2,2,2,1,1,4,2,5,0,0,0,0,0,5,5,0,5,0,0,0,0,5,302,,,5,0,0,0,5,5,,,,3,,,,2,0,,,,1,5,3,0,1,21,,,1,3,SAN LEANDRO,CA,,US,7,2,0,2,2,1,1,0.669478
25,468,15076,13,8,242,18,3,4,58,D,1,2,7:20 AM,6:10 AM,70,1,1,,,1,,,1,2,1,2,1,4,1,4,2,4,4,4,4,4,4,4,4,4,0,5,6,6,6,4,611,,,5,6,6,5,5,5,1,,,5,,,,1,5,,,,8,3,5,5,2,,,,1,8,FAIRFIELD,CA,94534,US,5,1,4,1,1,1,1,0.720538
26,567,15083,12,38,1777,99,4,2,1,A,1,3,8:00 AM,5:40 AM,140,1,2,,,2,,,,1,1,0,3,2,1,3,2,4,2,6,4,4,4,6,6,6,4,6,6,6,6,3,,,,5,6,6,2,3,3,,,,3,9,,,2,0,,,,6,5,5,4,1,1,,,1,2,REDWOOD CITY,CA,94061,US,3,1,1,3,2,3,2,0.669478
27,568,15083,12,38,1777,99,4,2,1,A,1,3,8:00 AM,6:00 AM,120,1,2,3,,2,,,,1,2,2,1,2,1,2,2,4,3,3,4,4,4,6,6,4,4,6,6,6,6,4,,,,5,6,6,6,4,4,,,,3,5,,,2,0,,,,1,5,2,5,2,,,,1,3,OAKLAND,CA,94621,US,2,2,1,2,1,0,1,0.669478
28,569,15083,12,38,1777,99,4,2,1,A,1,3,8:00 AM,5:30 AM,150,1,2,,,13,,,,1,1,1,2,1,1,4,2,5,4,4,5,5,5,4,5,6,5,5,6,6,6,5,,,,5,6,6,5,5,5,,,,5,,,,2,0,,,,3,5,5,5,2,,,,1,6,SAN RAFAEL,CA,94903,US,7,2,4,2,2,1,1,0.669478
29,570,15083,12,38,1777,99,4,2,1,A,1,3,8:00 AM,5:00 AM,180,1,1,,,6,,,,1,2,1,1,1,1,2,2,5,4,4,4,4,4,4,4,4,6,6,5,6,6,4,,,,5,5,5,5,5,5,,,,5,,,,2,0,,,,5,3,3,5,1,3,,,3,13,REGINA,SK,,CANADA,3,2,2,2,0,1,1,0.669478
30,469,15076,13,8,242,18,3,4,58,D,1,2,7:20 AM,5:30 AM,110,1,1,,,1,,,3,2,1,1,1,4,1,2,2,4,5,4,4,4,5,5,5,5,3,4,5,5,6,4,,,,5,5,6,5,4,5,,,,4,,,,2,0,,,,7,3,4,4,2,,,,1,4,SAN JOSE,CA,95112,US,5,1,3,2,1,2,1,0.720538
31,571,15083,12,38,1777,99,4,2,1,A,1,3,8:00 AM,5:30 AM,150,1,6,,,1,,,,2,1,1,0,0,3,1,4,2,5,5,5,5,5,5,5,5,5,5,5,5,5,5,999,,,5,5,5,5,5,5,,,,5,,,,1,5,,,,13,5,3,5,2,,,,2,10,LATHROP,CA,95330,US,5,2,4,2,1,1,1,0.669478
32,572,15083,12,38,1777,99,4,2,1,A,1,3,6:20 AM,N,N,1,4,,,2,,,,2,2,2,1,1,2,0,2,4,3,3,4,3,4,3,4,3,3,4,0,3,2,3,999,,,5,5,5,5,5,5,,,,5,5,,,0,0,,,,2,3,3,3,2,,,,0,99,0,0,0,0,0,0,0,0,0,0,1,0.669478
33,470,15076,13,8,242,18,3,4,58,D,1,2,7:20 AM,6:05 AM,75,1,3,,,2,,,,2,1,2,1,3,1,3,2,4,6,4,3,4,4,6,6,4,4,6,4,6,6,4,8,,,4,4,6,6,4,4,,,,5,5,,,1,5,1,,,7,5,4,5,2,,,,2,10,SANTA CRUZ,CA,95062,US,7,2,1,2,1,2,1,0.720538
34,573,15083,12,38,1777,99,4,2,1,A,1,3,8:00 AM,6:00 AM,120,1,2,,,2,,,,1,0,1,1,1,2,1,2,1,1,2,4,5,4,5,5,5,4,4,4,4,4,4,208,506,,0,0,0,3,0,0,,,,2,41,,,2,0,,,,0,0,5,0,2,,,,3,13,EDMONTON,AB,,CANADA,4,1,2,1,0,0,1,0.669478
35,574,15083,12,38,1777,99,4,2,1,A,1,3,8:00 AM,6:30 AM,90,1,2,3,,2,,,,1,1,1,1,3,1,3,2,5,3,3,4,5,5,2,2,5,5,3,5,5,5,4,102,,,5,3,3,3,2,3,,,,3,98,,,1,3,7,,,1,3,4,2,2,,,,3,13,VANCOUVER,BC,,CANADA,3,1,2,1,2,1,1,0.669478
36,471,15076,13,8,242,18,3,4,58,D,1,2,7:20 AM,6:15 AM,65,1,2,3,,2,,,,2,2,1,2,1,2,1,2,5,4,4,4,4,4,4,4,6,5,6,6,6,5,24,,,5,6,6,5,5,5,,,,5,,,,1,5,1,,,3,3,5,5,2,,,,3,11,KINGSTON,IL,60145,US,6,2,2,2,2,2,1,0.720538
37,575,15083,12,38,1777,99,4,2,1,A,1,3,8:00 AM,6:50 AM,70,1,6,,,2,,,,2,2,1,2,4,1,4,2,3,2,2,4,4,4,4,4,6,3,3,3,4,6,4,,,,5,6,6,4,4,4,,,,5,,,,2,0,,,,7,3,4,4,2,,,,1,4,SUNNYVALE,CA,94087,US,4,1,3,2,1,2,1,0.669478
38,576,15083,12,38,1777,99,4,2,1,A,1,3,8:00 AM,6:45 AM,75,1,3,,,6,,,,2,1,2,0,3,1,1,2,3,3,3,3,4,4,6,6,1,6,6,2,6,6,3,507,,,5,5,6,5,4,4,,,,5,5,,,2,0,,,,1,3,3,5,1,9,,,3,13,VANCOUVER,BC,,CANADA,3,2,2,3,0,1,1,0.669478
39,472,15076,13,8,242,18,3,4,58,D,1,2,7:20 AM,7:00 AM,20,1,3,,,2,,,,2,2,2,2,1,1,4,2,5,4,4,5,5,4,4,6,6,5,4,5,6,6,5,,,,5,5,6,5,5,5,,,,5,1,21,,2,0,,,,5,5,5,5,2,,,,1,1,SAN FRANCISC,CA,94119,US,6,2,0,2,2,2,1,0.720538
40,577,15083,12,38,1777,99,4,2,1,A,1,3,8:00 AM,4:08 AM,232,1,0,,,2,,,,1,2,1,1,1,1,3,2,6,3,2,3,4,5,2,2,2,5,5,6,6,6,3,704,701,705,5,6,6,4,5,5,,,,5,5,,,2,0,,,,1,5,5,5,2,,,,3,13,TORONTO,ON,,CANADA,4,2,0,3,1,2,1,0.669478
41,473,15076,13,8,242,18,3,4,58,D,1,2,7:20 AM,5:45 AM,95,1,6,,,2,,,,2,2,1,1,2,1,4,2,3,4,4,4,4,3,6,6,4,5,5,6,6,4,4,,,,4,6,3,4,4,4,,,,4,4,,,1,5,2,,,5,4,5,5,2,,,,3,12,,NY,13601,US,4,1,3,2,2,2,1,0.720538
42,474,15076,13,8,242,18,3,4,58,D,1,2,7:20 AM,6:30 AM,50,1,2,,,2,,,,2,2,2,1,3,1,3,2,5,5,5,5,5,5,6,5,4,4,4,6,4,6,5,,,,5,6,6,6,5,5,,,,5,,,,2,0,,,,1,4,5,0,2,,,,1,90,,CA,,US,5,2,4,2,2,1,1,0.720538
43,513,15079,12,8,5988,50,1,4,58,D,1,1,10:53 AM,10:00 AM,53,1,1,,,2,,,,1,1,2,2,3,1,4,1,6,6,6,5,6,5,6,6,6,5,6,6,6,6,5,8,,,5,6,6,6,5,5,,,,5,1,,,1,5,1,,,5,5,5,5,2,,,,3,10,SAN DIEGO,CA,92056,US,7,2,0,2,1,1,1,0.720538
44,475,15076,13,8,242,18,3,4,58,D,1,2,7:20 AM,6:00 AM,80,1,1,,,4,,,,2,1,1,2,1,1,2,2,3,4,5,5,5,5,6,6,6,6,6,6,6,6,4,,,,4,6,6,5,5,5,,,,5,,,,2,0,,,,5,5,5,4,2,,,,3,12,ROCHESTER,NY,14586,US,4,1,1,2,2,2,1,0.720538
45,514,15079,12,8,5988,50,1,4,58,D,1,1,10:53 AM,8:30 AM,143,1,4,,,2,,,,1,1,1,1,4,1,4,1,4,4,4,3,4,4,3,3,5,4,3,3,3,3,4,113,,,4,3,3,4,4,4,,,,2,41,,,2,0,,,,17,2,4,1,2,,,,2,10,SANTA CRUZ,CA,95060,US,3,1,4,1,1,1,1,0.720538
46,476,15076,13,8,242,18,3,4,58,D,1,2,7:20 AM,5:30 AM,110,1,2,,,2,,,,1,1,1,2,2,1,1,2,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,316,,,5,5,5,5,5,5,,,,5,,,,2,0,,,,2,5,3,5,1,1,,,1,5,PLEASANT HIL,CA,94523,US,3,2,1,2,2,2,1,0.720538
47,515,15079,12,8,5988,50,1,4,58,D,1,1,10:53 AM,9:00 AM,113,1,3,,,6,,,,2,2,1,1,2,1,3,2,6,5,5,5,5,5,6,6,5,5,6,6,6,6,5,,,,5,6,6,5,5,5,,,,5,,,,2,0,,,,1,5,5,5,2,,,,1,3,OAKLAND,CA,91786,US,3,2,1,2,1,0,1,0.720538
48,477,15076,13,8,242,18,3,4,58,D,1,2,7:20 AM,6:30 AM,50,1,1,,,1,,,0,2,0,0,0,3,1,4,2,4,5,4,4,4,4,4,4,4,3,3,3,2,6,4,,,,5,4,6,3,3,4,,,,4,,,,0,0,,,,3,4,5,5,2,,,,1,6,KENTFIELD,CA,94904,US,4,2,4,2,2,1,1,0.720538
49,516,15079,12,8,5988,50,1,4,58,D,1,1,10:53 AM,9:30 AM,83,1,1,,,5,,,,2,2,2,2,4,1,4,2,5,4,4,5,5,4,6,6,4,5,3,4,6,6,5,205,,,5,4,5,5,5,5,,,,5,,,,1,5,1,,,1,2,5,5,2,,,,1,3,ALAMEDA,CA,94502,US,5,2,4,1,2,1,1,0.720538
50,517,15079,12,8,5988,50,1,4,58,D,1,1,10:53 AM,9:15 AM,98,1,1,,,12,,,,2,2,1,2,2,1,4,2,3,3,3,3,3,4,6,5,6,4,4,6,6,6,4,,,,4,6,6,4,4,4,,,,5,6,27,,2,0,,,,5,3,4,5,2,,,,3,10,RESEDA,CA,91335,US,4,2,4,2,2,2,1,0.720538
51,518,15079,12,8,5988,50,1,4,58,D,1,1,10:53 AM,9:10 AM,103,1,4,,,13,,,,2,1,2,2,3,1,3,2,5,4,4,5,5,5,6,6,5,5,6,6,6,6,5,,,,5,6,6,5,5,5,,,,5,,,,2,0,,,,3,5,5,5,2,,,,3,10,LOS ANGELES,CA,91307,US,2,2,2,2,2,2,1,0.720538
52,478,15076,13,8,242,18,3,4,58,D,1,2,7:20 AM,6:40 AM,40,1,1,,,4,,,,0,1,0,1,6,1,1,2,5,5,5,5,5,5,6,5,5,5,4,5,6,4,5,108,,,5,5,3,4,4,5,1,,,5,4,,,1,5,11,9,,5,5,5,5,2,,,,1,1,SAN FRANCISC,CA,94109,US,5,2,5,1,2,2,1,0.720538
53,519,15079,12,8,5988,50,1,4,58,D,1,1,10:53 AM,9:00 AM,113,1,1,,,5,,,,1,2,2,2,4,1,4,2,4,4,3,4,4,4,3,4,3,3,3,4,6,6,4,,,,4,4,6,4,4,4,,,,4,,,,4,6,,,,5,2,3,3,2,,,,1,1,SAN FRANCISC,CA,,US,4,2,3,2,1,1,1,0.720538
54,520,15079,12,8,5988,50,1,4,58,D,1,1,10:53 AM,9:30 AM,83,1,1,,,5,,,,2,2,1,1,3,1,1,2,3,3,3,3,3,3,3,3,2,2,2,2,2,2,1,,,,4,4,4,4,4,4,,,,2,,,,2,0,,,,5,5,5,5,2,,,,0,99,0,0,0,0,0,0,0,0,0,0,1,0.720538
55,479,15076,13,8,242,18,3,4,58,D,1,2,7:20 AM,6:00 AM,80,1,2,,,6,,,,0,0,1,0,1,1,4,2,3,4,3,4,4,4,6,6,6,6,6,6,6,6,4,,,,4,6,6,4,4,4,,,,5,,,,1,5,,,,5,5,5,5,2,,,,3,12,BOSTON,MA,1945,US,7,2,2,2,2,2,1,0.720538
56,521,15079,12,8,5988,50,1,4,58,D,1,1,10:53 AM,9:30 AM,83,1,3,,,2,,,,2,2,2,2,3,1,2,2,4,4,4,4,4,5,4,4,5,5,6,5,6,4,5,,,,4,5,0,4,4,4,,,,4,,,,1,5,,,,5,4,4,5,1,3,,,1,1,SAN FRANCISC,CA,,US,7,2,0,2,2,2,1,0.720538
```

# Hoja de Calculo

- SFO Airport Survey Results

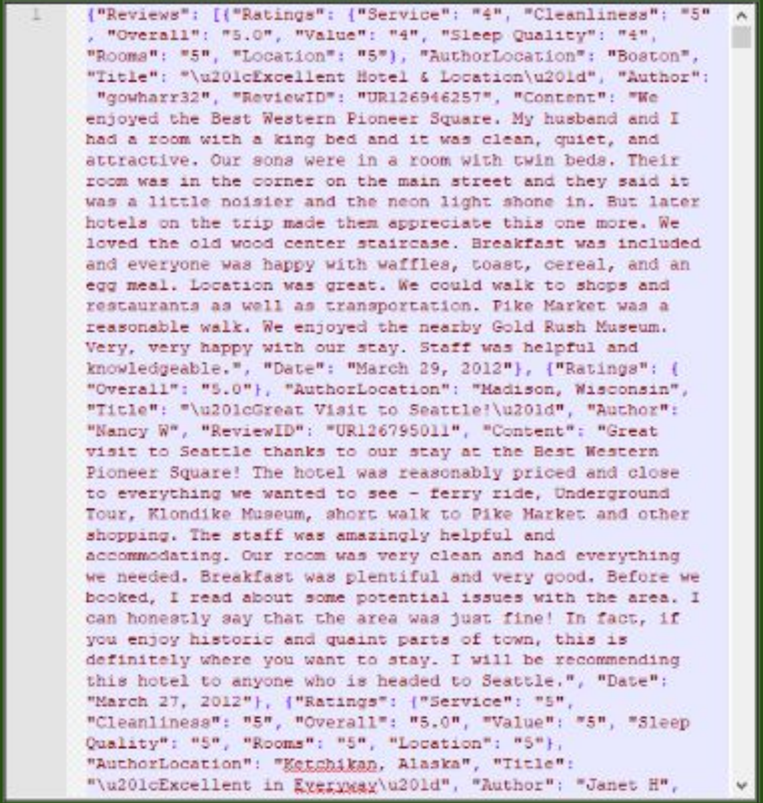


M1 | *fx* DEST

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | RESPNU | CCGI | RUN | INTDAT | GAT | ARE | STRAT | PEA | METHC | SA | AIRLIN | FLIGH | DES | DESTGE | DESTMAF | DEPTIM | ARRTIM | HOWL |
| 2 | 1 | 348 | 18045 | 4 | 54 | D | 3 | 2 | 1 | 1 | 36 | 182 | 62 | 2 | 3 | 9:20 PM | 6:30 PM | 170 |
| 3 | 2 | 349 | 18045 | 4 | 54 | D | 3 | 2 | 1 | 1 | 36 | 182 | 62 | 2 | 3 | 9:20 PM | 7:00 PM | 140 |
| 4 | 3 | 350 | 18045 | 4 | 54 | D | 3 | 2 | 1 | 1 | 36 | 182 | 62 | 2 | 3 | 9:20 PM | 8:00 PM | 80 |
| 5 | 4 | 351 | 18045 | 4 | 54 | D | 3 | 2 | 1 | 2 | 36 | 182 | 62 | 2 | 3 | 9:20 PM | 8:10 PM | 70 |
| 6 | 5 | 352 | 18045 | 4 | 54 | D | 3 | 2 | 1 | 2 | 36 | 182 | 62 | 2 | 3 | 9:20 PM | 8:20 PM | 60 |
| 7 | 6 | 353 | 18045 | 4 | 54 | D | 3 | 2 | 1 | 2 | 36 | 182 | 62 | 2 | 3 | 9:20 PM | 7:00 PM | 140 |
| 8 | 7 | 354 | 18045 | 4 | 54 | D | 3 | 2 | 1 | 2 | 36 | 182 | 62 | 2 | 3 | 9:20 PM | 7:00 PM | 140 |
| 9 | 8 | 355 | 18045 | 4 | 54 | D | 3 | 2 | 1 | 2 | 36 | 182 | 62 | 2 | 3 | 9:20 PM | 8:07 PM | 73 |
| 10 | 9 | 356 | 18045 | 4 | 54 | D | 3 | 2 | 1 | 2 | 36 | 182 | 62 | 2 | 3 | 9:20 PM | 6:00 PM | 200 |
| 11 | 10 | 357 | 18045 | 4 | 54 | D | 3 | 2 | 1 | 2 | 36 | 182 | 62 | 2 | 3 | 9:20 PM | 1:00 PM | 500 |
| 12 | 11 | 358 | 18045 | 4 | 54 | D | 3 | 2 | 1 | 2 | 36 | 182 | 62 | 2 | 3 | 9:20 PM | 7:50 PM | 90 |
| 13 | 12 | 359 | 18045 | 4 | 54 | D | 3 | 2 | 1 | 2 | 36 | 182 | 62 | 2 | 3 | 9:20 PM | 6:00 PM | 200 |
| 14 | 13 | 360 | 18045 | 4 | 54 | D | 3 | 2 | 1 | 2 | 36 | 182 | 62 | 2 | 3 | 9:20 PM | 7:20 PM | 120 |
| 15 | 14 | 361 | 18045 | 4 | 54 | D | 3 | 2 | 1 | 2 | 36 | 182 | 62 | 2 | 3 | 9:20 PM | 8:00 PM | 80 |
| 16 | 15 | 362 | 18045 | 4 | 54 | D | 3 | 2 | 1 | 2 | 36 | 182 | 62 | 2 | 3 | 9:20 PM | 7:30 PM | 110 |
| 17 | 16 | 363 | 18045 | 4 | 54 | D | 3 | 2 | 1 | 2 | 36 | 182 | 62 | 2 | 3 | 9:20 PM | 7:30 PM | 110 |
| 18 | 17 | 364 | 18045 | 4 | 54 | D | 3 | 2 | 1 | 2 | 36 | 182 | 62 | 2 | 3 | 9:20 PM | 6:30 PM | 170 |
| 19 | 18 | 365 | 18045 | 4 | 54 | D | 3 | 2 | 1 | 2 | 36 | 182 | 62 | 2 | 3 | 9:20 PM | 7:20 PM | 120 |
| 20 | 19 | 366 | 18045 | 4 | 54 | D | 3 | 2 | 1 | 2 | 36 | 182 | 62 | 2 | 3 | 9:20 PM | 7:30 PM | 110 |
| 21 | 20 | 367 | 18045 | 4 | 54 | D | 3 | 2 | 1 | 2 | 36 | 182 | 62 | 2 | 3 | 9:20 PM | 6:30 PM | 170 |

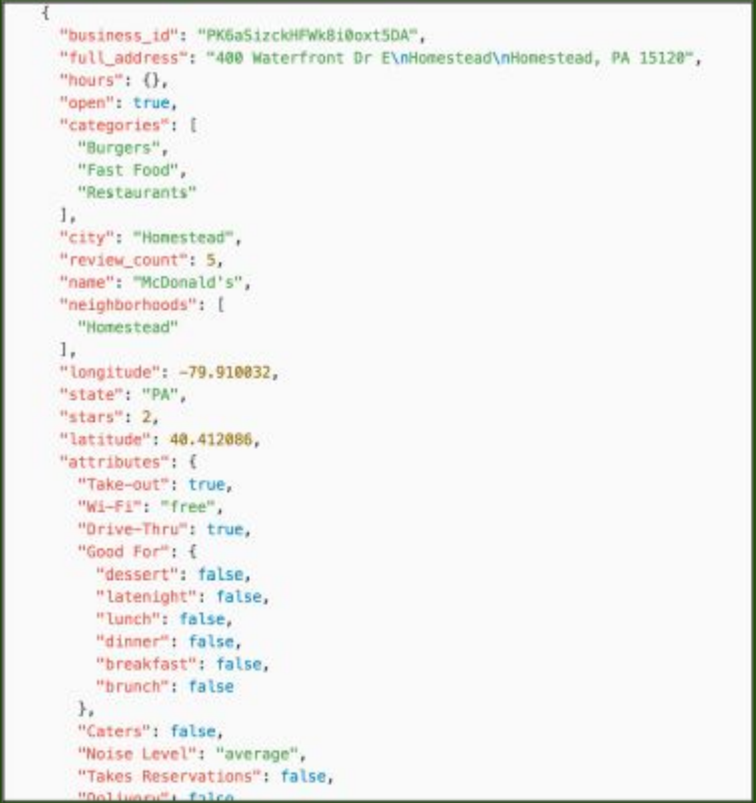
# JSON Examples

- Trip Advisor JSON File



```
{"Reviews": [{"Ratings": {"Service": "4", "Cleanliness": "5"
, "Overall": "5.0", "Value": "4", "Sleep Quality": "4",
"Rooms": "5", "Location": "5"}, "AuthorLocation": "Boston",
"Title": "\u201cExcellent Hotel & Location\u201d", "Author":
 "gowharr32", "ReviewID": "UR126946257", "Content": "We
enjoyed the Best Western Pioneer Square. My husband and I
had a room with a king bed and it was clean, quiet, and
attractive. Our sons were in a room with twin beds. Their
room was in the corner on the main street and they said it
was a little noisier and the neon light shone in. But later
hotels on the trip made them appreciate this one more. We
loved the old wood center staircase. Breakfast was included
and everyone was happy with waffles, toast, cereal, and an
egg meal. Location was great. We could walk to shops and
restaurants as well as transportation. Pike Market was a
reasonable walk. We enjoyed the nearby Gold Rush Museum.
Very, very happy with our stay. Staff was helpful and
knowledgeable.", "Date": "March 29, 2012"}, {"Ratings": {
"Overall": "5.0"}, "AuthorLocation": "Madison, Wisconsin",
"Title": "\u201cGreat Visit to Seattle!\u201d", "Author":
"Nancy W", "ReviewID": "UR126795011", "Content": "Great
visit to Seattle thanks to our stay at the Best Western
Pioneer Square! The hotel was reasonably priced and close
to everything we wanted to see - ferry ride, Underground
Tour, Klondike Museum, short walk to Pike Market and other
shopping. The staff was amazingly helpful and
accommodating. Our room was very clean and had everything
we needed. Breakfast was plentiful and very good. Before we
booked, I read about some potential issues with the area. I
can honestly say that the area was just fine! In fact, if
you enjoy historic and quaint parts of town, this is
definitely where you want to stay. I will be recommending
this hotel to anyone who is headed to Seattle.", "Date":
"March 27, 2012"}, {"Ratings": {"Service": "5",
"Cleanliness": "5", "Overall": "5.0", "Value": "5", "Sleep
Quality": "5", "Rooms": "5", "Location": "5"},
"AuthorLocation": "Ketchikan, Alaska", "Title":
"\u201cExcellent in Everyway\u201d", "Author": "Janet H",
```

* Yelp JSON file



```
{
  "business_id": "PK6aSizckHFWk8i0oxt5DA",
  "full_address": "400 Waterfront Dr E\nHomestead\nHomestead, PA 15120",
  "hours": {},
  "open": true,
  "categories": [
    "Burgers",
    "Fast Food",
    "Restaurants"
  ],
  "city": "Homestead",
  "review_count": 5,
  "name": "McDonald's",
  "neighborhoods": [
    "Homestead"
  ],
  "longitude": -79.910032,
  "state": "PA",
  "stars": 2,
  "latitude": 40.412086,
  "attributes": {
    "Take-out": true,
    "Wi-Fi": "free",
    "Drive-Thru": true,
    "Good For": {
      "dessert": false,
      "latenight": false,
      "lunch": false,
      "dinner": false,
      "breakfast": false,
      "brunch": false
    },
    "Caters": false,
    "Noise Level": "average",
    "Takes Reservations": false,
    "Delivery": false
```

# SQL Examples

- Sanoke customer data

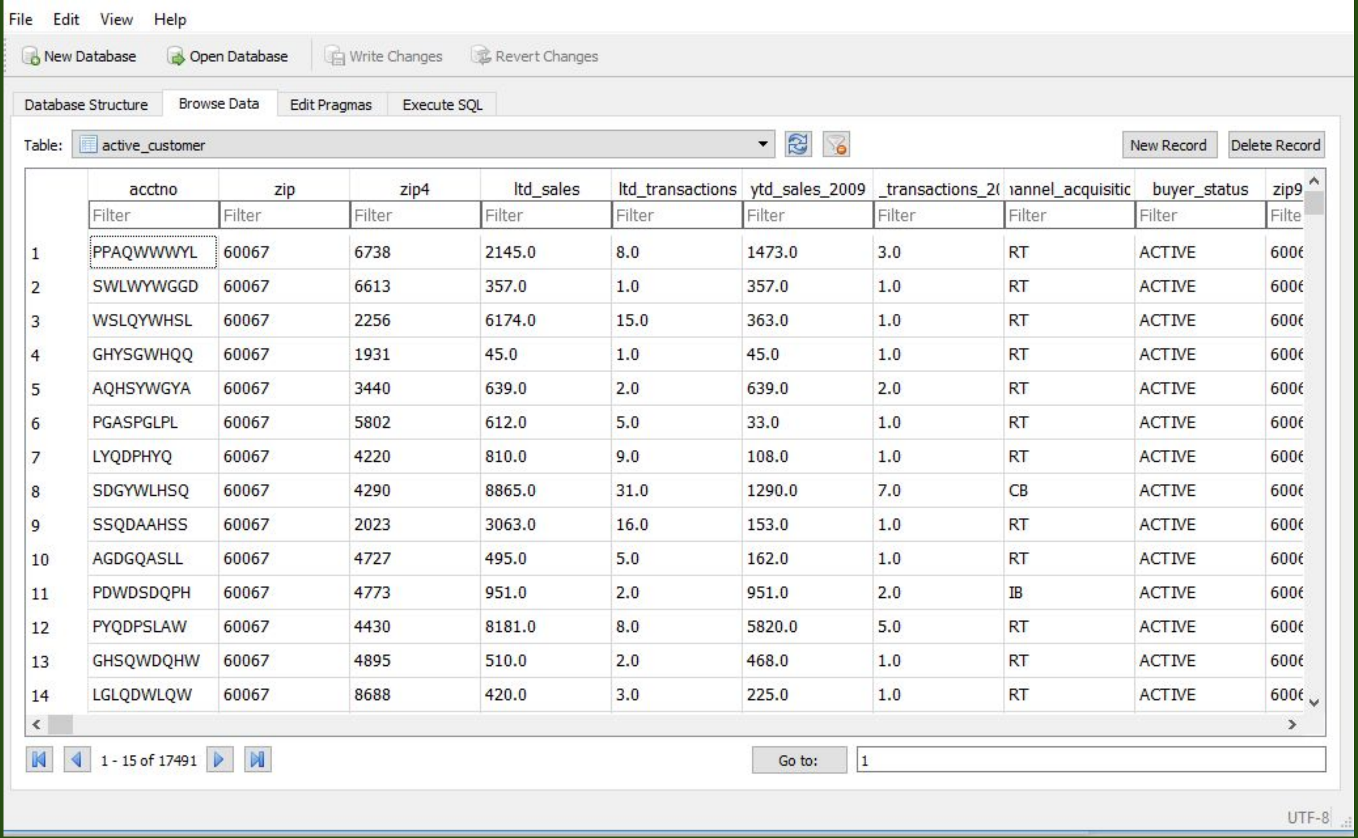


Table: active_customer

| | acctno | zip | zip4 | ltd_sales | ltd_transactions | ytd_sales_2009 | _transactions_2( | ıannel_acquisitic | buyer_status | zip9 |
|---|---|---|---|---|---|---|---|---|---|---|
| | Filter | Filter | Filter | Filter | Filter | Filter | Filter | Filter | Filter | Filte |
| 1 | PPAQWWWYL | 60067 | 6738 | 2145.0 | 8.0 | 1473.0 | 3.0 | RT | ACTIVE | 6006 |
| 2 | SWLWYWGGD | 60067 | 6613 | 357.0 | 1.0 | 357.0 | 1.0 | RT | ACTIVE | 6006 |
| 3 | WSLQYWHSL | 60067 | 2256 | 6174.0 | 15.0 | 363.0 | 1.0 | RT | ACTIVE | 6006 |
| 4 | GHYSGWHQQ | 60067 | 1931 | 45.0 | 1.0 | 45.0 | 1.0 | RT | ACTIVE | 6006 |
| 5 | AQHSYWGYA | 60067 | 3440 | 639.0 | 2.0 | 639.0 | 2.0 | RT | ACTIVE | 6006 |
| 6 | PGASPGLPL | 60067 | 5802 | 612.0 | 5.0 | 33.0 | 1.0 | RT | ACTIVE | 6006 |
| 7 | LYQDPHYQ | 60067 | 4220 | 810.0 | 9.0 | 108.0 | 1.0 | RT | ACTIVE | 6006 |
| 8 | SDGYWLHSQ | 60067 | 4290 | 8865.0 | 31.0 | 1290.0 | 7.0 | CB | ACTIVE | 6006 |
| 9 | SSQDAAHSS | 60067 | 2023 | 3063.0 | 16.0 | 153.0 | 1.0 | RT | ACTIVE | 6006 |
| 10 | AGDGQASLL | 60067 | 4727 | 495.0 | 5.0 | 162.0 | 1.0 | RT | ACTIVE | 6006 |
| 11 | PDWDSDQPH | 60067 | 4773 | 951.0 | 2.0 | 951.0 | 2.0 | IB | ACTIVE | 6006 |
| 12 | PYQDPSLAW | 60067 | 4430 | 8181.0 | 8.0 | 5820.0 | 5.0 | RT | ACTIVE | 6006 |
| 13 | GHSQWDQHW | 60067 | 4895 | 510.0 | 2.0 | 468.0 | 1.0 | RT | ACTIVE | 6006 |
| 14 | LGLQDWLQW | 60067 | 8688 | 420.0 | 3.0 | 225.0 | 1.0 | RT | ACTIVE | 6006 |

1 - 15 of 17491

Go to: 1

# Python

- Object-oriented, high-level programming  language
- Used as a scripting or glue language to connect existing components together
- Simple, easy to learn syntax emphasizes readability
- Supports modules and packages
- Python interpreter and the extensive standard library are FREE!

# Python

Key Python Package:

- Pandas
    - Open source library that allows you to work with CSV, Excel, JSON, and SQL database files, pull them into tables (called dataframes), and perform various data analysis techniques.
-

# References

- ★ Kernighan, Brian W., and Dennis M. Ritchie. The C Programming Language. Vol. 2. Englewood Cliffs: prentice-Hall, 1988.
- ★ Silberschatz, Abraham, Peter B. Galvin, and Greg Gagne. Operating System Concepts. Vol. 8. Wiley, 2013.
- ★ https://planningtank.com/computer-applications/data-processing-cycle
- ★ https://www.talend.com/resources/what-is-data-processing/
- ★ https://peda.net/kenya/ass/subjects2/computer-studies/form-3/data-processing/dpc2
- ★ https://www.academia.edu/38210518/What_is_Data_Processing
- ★ https://www.studocu.com/en/document/polytechnic-university-of-the-philippines/information-and-communication-technology/lecture-notes/data-processing-lectures-in-data-processing/3167716/view
- ★ http://download.nos.org/srsec330/330L2.pdf