Lead Scoring Business Analytics Tools

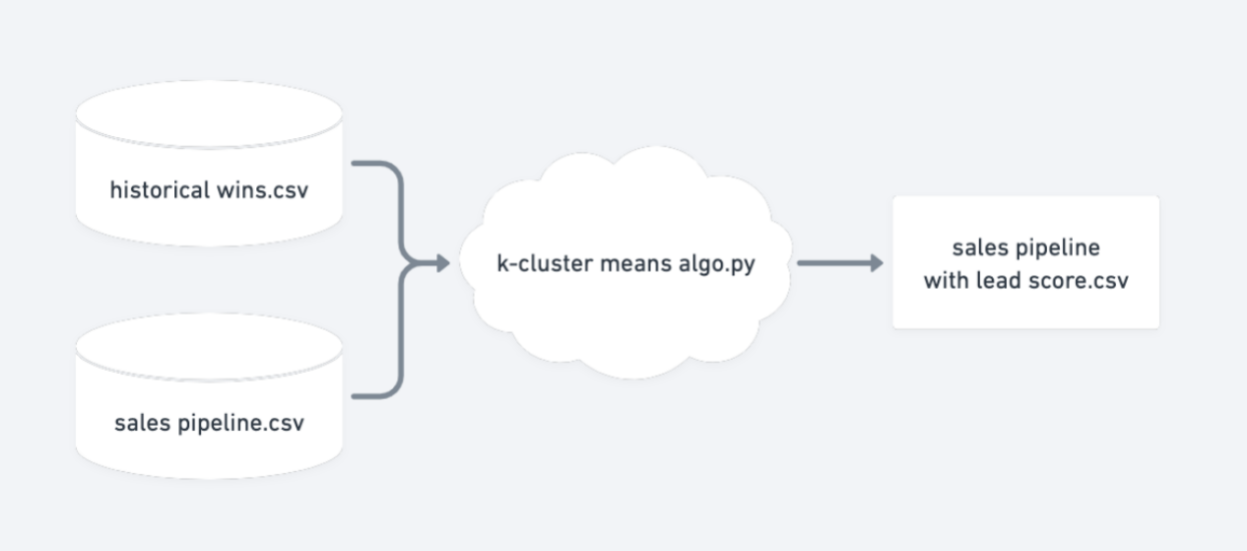
About

Python program designed to cluster historical wins (weddings business) into 5 different customer segments. Each customer is segmented using three data points:

1. Budget
2. Number of wedding guests
3. Days until wedding

Once historical wins customer segments are established, the program inserts a lead, one at a time, and reruns the clustering algorithm to group the lead into one of five clusters. Leads are then printed on a new excel spreadsheet with their corresponding Deal ID, parameters, and cluster number.

Program Flow



Program takes historical wins spreadsheet and sales pipeline spreadsheet as inputs, and outputs a third spreadsheet with corresponding clusters.

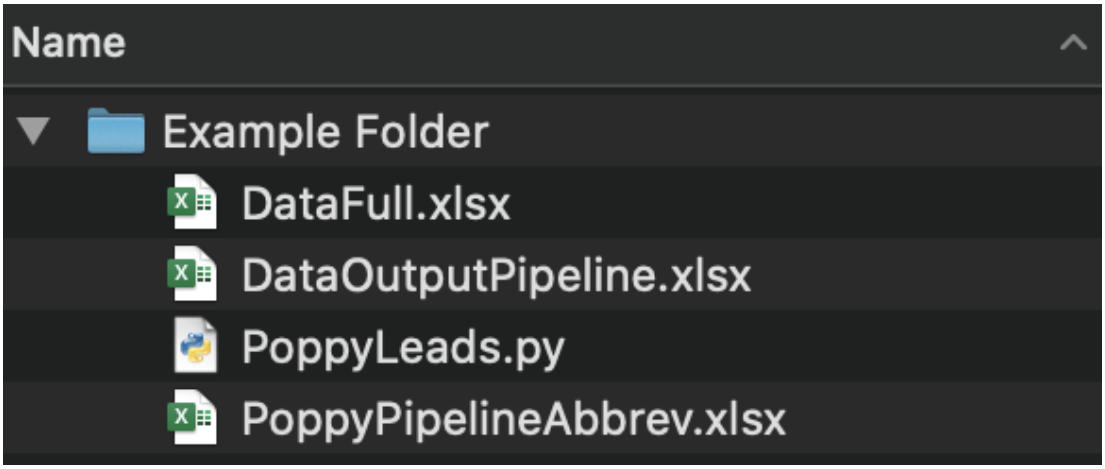
Interpreting Clusters

```
kevinjennings@Kevins-MacBook-Pro Poppy % Python3.9 PoppyPython.py
     Budget  Guest_Count  Days_From_Lead_To_Event  Cluster
0  0.929829     0.943727                 0.877034        0
1 -0.923736    -1.050654                -0.983293        1
2 -0.914198    -0.205062                 0.794382        2
3  0.364715     0.744987                -0.791146        3
4  0.730600    -0.745786                 0.416367        4
```

In the above text box, the cluster number is y-axis and the customer parameters are the x-axis. The numbers corresponding to each cell are the average number corresponding to each cluster, with an average of 1 and standard deviation of 0 for each. To further clarify, each number is on a scale from -1 to 1. Given this, segment 0 represents a customer segment with the highest budget, guest count, and days to the wedding. Segment 4 represents a customer segment with a high willingness to pay for a smaller guest count.

Instructions for Running

1. To run the program, organize the 3 spreadsheets and the python program into the same folder:



2. Navigate to the folder in terminal (if using a Mac).

```
kevinjennings@Kevins-MacBook-Pro Example Folder % ls
DataFull.xlsx                   PoppyLeads.py
DataOutputPipeline.xlsx         PoppyPipelineAbbrev.xlsx
kevinjennings@Kevins-MacBook-Pro Example Folder % pwd
/Users/kevinjennings/Documents/Georgetown/InSITE/Example Folder
```

3. Run the program using the following command:

```
kevinjennings@Kevins-MacBook-Pro Example Folder % Python3 PoppyLeads.py
```

4. Leads with clusters are now populated in “DataOutputPipeline.xlsx”

	A	B	C	D	E	F	G
1	Deal_ID	Budget	Guest_Count	Days_From_Lead_To_Event	Cluster		
54	3022	2000	150	342	3		
55	3042	1000	100	266	2		
56	3050	3000	150	262	3		
57	3051	1000	75	156	0		
58	3053	1000	50	120	0		
59	3059	1000	20	63	0		
60	3074	1000	20	23	0		
61	3083	2000	75	221	1		
62	3090	1000	50	136	0		
63	3145	1000	0	17	0		
64	3149	1000	150	646	2		
65	3160	2500	100	331	1		
66	3187	2000	75	292	1		
67	3189	3000	150	251	3		
68	3206	1000	20	228	2		
69	3212	1000	100	255	2		
70	3241	1000	20	97	0		
71	3253	2000	150	222	3		
72	3260	2000	20	137	1		