Week 8 Lab - Visualizations

This week's assignment will focus on using Tableau to produce insights and visualizations for a dataset of your choice.

Data:

Find an "interesting" data set to work with. UCI Machine Learning Archive and data.gov are always good places to start. You will need two datasets in total.

Task 1:

For one of your chosen datasets, complete the following

1. Create at least 3 descriptive graphics to communicate some aspect of your data.
2. Pay attention to the following:
    - Use of colors
    - Labeling of your axes
    - Descriptive title for your charts
3. Make sure you use the following at least once within your 3 graphics
    - Change the Tableau default data display type (Think back to our Date field being aggregated by year)
    - Measure Values
    - Calculated field
4. Create at least one dashboard
5. Create a Story

Task 2:

For your second chosen datasets, complete the following.

1. Create at least one graphic that was not demostrated in the Lecture FTE. Take a look at this Tableau Tutorial page for additional graphic types.
2. Create a dashboard with your graphic(s) from step 1.

Important: Make sure your graphics are fairly self explanatory. If you need to provide additional explanations of the information you are conveying to me, add this information to you Dashboard or Story page.

Deliverables:

Upload your Tableau workbooks to WorldClass. You can choice to do both Tasks in one workbook or separate workbooks.

I. Introduction

In this assignment, I was tasked to find 2 datasets and use Tableau to find relationships between the data and tell a story. I found that much more time was needed learn Tableau in order to properly complete this assignment. Working with a brand new software and 2 brand new datasets proved more work than anticipated! This was a very challenging yet interesting assignment as I'm still very new to Tableau and trying to do my best to learn the new software.

II. Methods, III. Code, and IV. Analysis of Results

In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

import warnings
warnings.filterwarnings("ignore")

%matplotlib inline
sns.set()
```

In [2]:
```
df1 = pd.read_csv("data-scientist-salaries/data_cleaned_2021.csv")
df1.head()
```

Out[2]:

	index	Job Title	Salary Estimate	Job Description	Rating	Company Name	Location	Headquarters	Size	Founded	...	tensor	hadoop	tableau	bi
0	0	Data Scientist	$53K-91K (Glassdoor est.)	Data Scientist\nLocation: Albuquerque, NM\nEdu...	3.8	Tecolote Research\n3.8	Albuquerque, NM	Goleta, CA	501 - 1000	1973	...	0	0	1	1
1	1	Healthcare Data Scientist	$63K-112K (Glassdoor est.)	What You Will Do:\n\nI. General Summary\n\nThe...	3.4	University of Maryland Medical System\n3.4	Linthicum, MD	Baltimore, MD	10000+	1984	...	0	0	0	0
2	2	Data Scientist	$80K-90K (Glassdoor est.)	KnowBe4, Inc. is a high growth information sec...	4.8	KnowBe4\n4.8	Clearwater, FL	Clearwater, FL	501 - 1000	2010	...	0	0	0	0

	index	Job Title	Salary Estimate	Job Description	Rating	Company Name	Location	Headquarters	Size	Founded	...	tensor	hadoop	tableau	bi
3	3	Data Scientist	$56K-$97K (Glassdoor est.)	*Organization and Job ID**\nJob ID: 310709\n\n...	3.8	PNNL\n3.8	Richland, WA	Richland, WA	1001 - 5000	1965	...	0	0	0	0
4	4	Data Scientist	$86K-$143K (Glassdoor est.)	Data Scientist\nAffinity Solutions / Marketing...	2.9	Affinity Solutions\n2.9	New York, NY	New York, NY	51 - 200	1998	...	0	0	0	0

5 rows × 42 columns

In [3]:
```
df1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 742 entries, 0 to 741
Data columns (total 42 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   index               742 non-null    int64
 1   Job Title           742 non-null    object
 2   Salary Estimate     742 non-null    object
 3   Job Description     742 non-null    object
 4   Rating              742 non-null    float64
 5   Company Name        742 non-null    object
 6   Location            742 non-null    object
 7   Headquarters        742 non-null    object
 8   Size                742 non-null    object
 9   Founded             742 non-null    int64
 10  Type of ownership   742 non-null    object
 11  Industry            742 non-null    object
 12  Sector              742 non-null    object
 13  Revenue             742 non-null    object
 14  Competitors         742 non-null    object
 15  Hourly              742 non-null    int64
 16  Employer provided   742 non-null    int64
 17  Lower Salary        742 non-null    int64
 18  Upper Salary        742 non-null    int64
 19  Avg Salary(K)       742 non-null    float64
 20  company_txt         742 non-null    object
 21  Job Location        742 non-null    object
 22  Age                 742 non-null    int64
 23  Python              742 non-null    int64
 24  spark               742 non-null    int64
 25  aws                 742 non-null    int64
```

```
 26  excel               742 non-null    int64
 27  sql                 742 non-null    int64
 28  sas                 742 non-null    int64
 29  keras               742 non-null    int64
 30  pytorch             742 non-null    int64
 31  scikit              742 non-null    int64
 32  tensor              742 non-null    int64
 33  hadoop              742 non-null    int64
 34  tableau             742 non-null    int64
 35  bi                  742 non-null    int64
 36  flink               742 non-null    int64
 37  mongo               742 non-null    int64
 38  google_an           742 non-null    int64
 39  job_title_sim       742 non-null    object
 40  seniority_by_title  742 non-null    object
 41  Degree              742 non-null    object
dtypes: float64(2), int64(23), object(17)
memory usage: 243.6+ KB
```

In [4]:
```
df1.shape
```

Out[4]:
```
(742, 42)
```

I think I will be using the following columns for my data analysis / useful charts:

- Job Title
- Location
- Size
- Founded
- Industry
- Lower Salary
- Upper Salary
- Avg Salary

In [15]:
```
pd.set_option("display.max_rows", None)
```

In [16]:
```
df1.isnull().sum()
```

Out[16]:
```
index                0
Job Title            0
Salary Estimate      0
Job Description      0
Rating               0
Company Name         0
```

```
Location              0
Headquarters          0
Size                  0
Founded               0
Type of ownership     0
Industry              0
Sector                0
Revenue               0
Competitors           0
Hourly                0
Employer provided     0
Lower Salary          0
Upper Salary          0
Avg Salary(K)         0
company_txt           0
Job Location          0
Age                   0
Python                0
spark                 0
aws                   0
excel                 0
sql                   0
sas                   0
keras                 0
pytorch               0
scikit                0
tensor                0
hadoop                0
tableau               0
bi                    0
flink                 0
mongo                 0
google_an             0
job_title_sim         0
seniority_by_title    0
Degree                0
dtype: int64
```

In [17]:

```
df1['Job Title'].value_counts()
```

Out[17]:

```
Data Scientist                 131
Data Engineer                   53
Senior Data Scientist           34
Data Analyst                    15
Senior Data Engineer            14
Senior Data Analyst             12
Lead Data Scientist              8
Marketing Data Analyst           6
Sr. Data Engineer                6
```

Machine Learning Engineer	5
Principal Data Scientist	5
R&D Specialist/ Food Scientist	4
Medical Laboratory Scientist	4
Research Scientist	4
Senior Research Scientist-Machine Learning	4
MED TECH/LAB SCIENTIST- SOUTH COASTAL LAB	4
Analytics Manager - Data Mart	4
Food Scientist - Developer	4
Staff Scientist-Downstream Process Development	4
Sr. Data Engineer - Contract-to-Hire (Java)	4
Sr. Scientist Method Development	3
Project Scientist - Auton Lab, Robotics Institute	3
Scientist, Molecular/Cellular Biologist	3
Associate Scientist, LC/MS Biologics	3
Revenue Analytics Manager	3
IT - Data Engineer II	3
Research Scientist, Immunology - Cancer Biology	3
Scientist - Biomarker and Flow Cytometry	3
Senior Scientist - Regulatory Submissions	3
ENVIRONMENTAL ENGINEER/SCIENTIST	3
Senior Data Science Systems Engineer	3
Senior Insurance Data Scientist	3
Associate Data Analyst- Graduate Development Program	3
Scientist - Analytical Services	3
Principal Scientist, Chemistry & Immunology	3
Director II, Data Science - GRM Actuarial	3
Scientist/Senior Scientist, Autoimmune	3
Clinical Data Analyst	3
Staff Data Engineer	3
Consultant - Analytics Consulting	3
Lead Data Engineer	3
Principal Scientist, Hematology	3
Staff Machine Learning Engineer	3
Software Engineer - Data Visualization	3
Senior Data Scientist - R&D Oncology	3
Principal Data Scientist (Computational Chemistry)	3
Senior Scientist (Neuroscience)	3
Machine Learning Engineer - Regulatory	3
Scientist, Bacteriology	3
Associate Director, Platform and DevOps- Data Engineering and Aritifical Intelligence	3
Clinical Laboratory Scientist	3
Information Security Data Analyst	3
Sr. Data Analyst	3
Data Science Manager	3
Scientist	3
Staff Scientist- Upstream PD	2
Sr Expert Data Science, Advanced Visual Analytics (Associate level)	2
Sr Data Analyst - IT	2
Associate Machine Learning Engineer / Data Scientist May 2020 Undergrad	2

Senior Scientist - Biostatistician	2
Senior Data Scientist Oncology	2
Associate Principal Scientist, Pharmacogenomics	2
Data Scientist - Systems Engineering	2
Data Engineer - Consultant (Charlotte Based)	2
Data Analyst 1, full-time contract worker for up to 12 months	2
Scientist, Immuno-Oncology	2
Products Data Analyst II	2
Lead Data Analyst	2
Data Science Engineer - Mobile	2
IT Associate Data Analyst	2
Scientist, Pharmacometrics	2
Business Data Analyst	2
Principal, Data Science - Advanced Analytics	2
Data Science Project Manager	2
Sr Scientist, Immuno-Oncology - Oncology	2
Data Analytics Project Manager	2
Scientist Manufacturing - Kentucky BioProcessing	2
Director - Data, Privacy and AI Governance	2
Staff BI and Data Engineer	2
Associate Data Engineer	2
Research Scientist – Security and Privacy	2
Market Data Analyst	2
Product Engineer – Data Science	2
Computational Chemist/Data Scientist	2
Director Data Science	2
Senior Research Analytical Scientist-Non-Targeted Analysis	2
Systems Engineer II - Data Analyst	2
Product Engineer – Spatial Data Science and Statistical Analysis	2
Managing Data Scientist/ML Engineer	2
Technology-Minded, Data Professional Opportunities	2
Salesforce Analytics Consultant	2
Scientist Manufacturing Pharma - Kentucky BioProcessing	2
Sr Data Engineer (Sr BI Developer)	2
Associate, Data Science, Internal Audit	2
Associate Environmental Scientist - Wildlife Biologist	2
Data Modeler - Data Solutions Engineer	2
Marketing Data Analyst, May 2020 Undergrad	2
Corporate Risk Data Analyst (SQL Based) - Milwaukee or	2
Senior Manager, Epidemiologic Data Scientist	2
Enterprise Architect, Data	2
Lead Big Data Engineer	2
Sr Software Engineer (Data Scientist)	2
Senior Scientist - Toxicologist - Product Integrity (Stewardship)	2
Sr. Data Engineer (ETL Developer)	2
Senior Data Scientist Artificial Intelligence	2
Analytics - Business Assurance Data Analyst	2
Associate Director/Director, Safety Scientist	2
Senior Scientist, Cell Pharmacology/Assay Development	2
Data Analyst Senior	2

Lead Data Engineer (Python)	2
Senior Data Scientist 4 Artificial Intelligence	2
Medical Lab Scientist - MLT	2
Senior Operations Data Analyst, Call Center Operations	2
Director II, Data Science - GRS Predictive Analytics	2
RESEARCH COMPUTER SCIENTIST - RESEARCH ENGINEER - SR. COMPUTER SCIENTIST - SOFTWARE DEVELOPMENT	2
Geospatial Software Developer and Data Scientist	2
Senior LiDAR Data Scientist	2
Big Data Engineer	2
Medical Lab Scientist	2
Senior Data & Machine Learning Scientist	2
Radar Data Analyst	2
VP, Data Science	2
Machine Learning Research Scientist	2
Excel / VBA / SQL Data Analyst	2
Sr. Data Scientist II	2
Sr. Scientist, Quantitative Translational Sciences	2
MED TECH/LAB SCIENTIST - LABORATORY	2
Principal Data Scientist with over 10 years experience	2
Principal Scientist - Immunologist	2
Sr. Scientist - Digital & Image Analysis/Computational Pathology	2
Digital Marketing & ECommerce Data Analyst	2
Analytics Manager	2
Scientist, Analytical Development	2
Risk and Analytics IT, Data Scientist	2
Senior Scientist - Neuroscience	2
Data Engineer 5 - Contract (Remote)	2
Staff Data Scientist	2
Data Scientist (Actuary, FSA or ASA)	2
Pricipal Scientist Molecular and cellular biologist	2
Sr. Data Scientist - Analytics, Personalized Healthcare (PHC)	2
BI & Platform Analytics Manager	2
PL Actuarial-Lead Data Scientist	2
PV Scientist	2
Data Scientist - Algorithms & Inference	2
Data Scientist - Quantitative	2
Data Scientist, Office of Data Science	2
College Hire - Data Scientist - Open to December 2019 Graduates	2
Senior Risk Data Scientist	2
Staff Data Scientist - Technology	2
Data Scientist / Machine Learning Expert	2
Clinical Data Scientist	2
Associate Data Analyst	2
Digital Health Data Scientist	2
Senior Data Scientist / Machine Learning	2
Data Scientist in Artificial Intelligence Early Career	2
Data Scientist - Health Data Analytics	2
Customer Data Scientist	2
Data Scientist SR	2
Data Scientist - Alpha Insights	1

Senior Data Scientist - Algorithms	1
Data Engineer - ETL	1
Data Modeler (Analytical Systems)	1
Data Science Analyst	1
Data Scientist in Translational Medicine	1
Data Analyst 2 (Missionary Department)	1
Supply Chain Data Analyst	1
Spectral Scientist/Engineer	1
Web Data Analyst	1
Data Scientist - Sales	1
Scientist I/II, Biology	1
Data Engineer I	1
Senior Data Scientist – Visualization, Novartis AI Innovation Lab	1
Product Manager/Data Evangelist	1
Insurance Financial Data Analyst	1
Senior Data Analyst/Scientist	1
Scientist – Cancer Discovery, Molecular Assay	1
Associate Scientist / Sr. Associate Scientist, Antibody Discovery	1
Data Architect / Data Modeler	1
Jr. Data Scientist	1
Data Scientist (Warehouse Automation)	1
Scientist - CVRM Metabolism - in vivo pharmacology	1
Sr. Data Engineer | Big Data SaaS Pipeline	1
Assistant Director/Director, Office of Data Science	1
Manager, Safety Scientist, Medical Safety & Risk Management	1
Software Engineer Staff Scientist: Human Language Technologies	1
Clinical Scientist, Clinical Development	1
Quality Control Scientist III- Analytical Development	1
Senior Engineer, Data Management Engineering	1
Principal Research Scientist/Team Lead, Medicinal Chemistry - Oncology	1
Senior Health Data Analyst, Star Ratings	1
Foundational Community Supports Data Analyst	1
Research Scientist, Machine Learning Department	1
Research Scientist / Principal Research Scientist - Multiphysical Systems	1
Data Analyst Chemist - Quality System Contractor	1
Research Scientist or Senior Research Scientist - Computer Vision	1
Senior Quantitative Analyst	1
Senior Formulations Scientist II	1
Director, Precision Medicine Clinical Biomarker Scientist	1
Associate Research Scientist I (Protein Expression and Production)	1
Software Engineer (Data Scientist/Software Engineer) - SISW - MG	1
Data Scientist Manager	1
Manager of Data Science	1
Data Engineering Analyst	1
Software Data Engineer - College	1
Sr. Scientist II	1
Data Analyst, Performance Partnership	1
Junior Data Analyst	1
Senior Data Scientist Statistics	1
Senior Spark Engineer (Data Science)	1

Senior Research Statistician- Data Scientist	1
Business Data Analyst, SQL	1
Medical Technologist / Clinical Laboratory Scientist	1
Associate Data Scientist/Computer Scientist	1
Business Intelligence Analyst / Developer	1
System and Data Analyst	1
Data & Analytics Consultant (NYC)	1
Big Data Engineer - Chicago - Future Opportunity	1
Survey Data Analyst	1
Lead Health Data Analyst - Front End	1
Healthcare Data Scientist	1
Customer Data Scientist/Sales Engineer	1
Data Operations Lead	1
RESEARCH SCIENTIST - BIOLOGICAL SAFETY	1
Principal Data Engineer, Data Platform & Insights	1
Senior Data Scientist: Causal & Predictive analytics AI Innovation Lab	1
Program/Data Analyst	1
SQL Data Engineer	1
Associate Scientist/Scientist, Process Analytical Technology - Small Molecule Analytical Chemistry	1
Staff Scientist	1
Data Engineer, Data Engineering and Artifical Intelligence	1
CONSULTANT– DATA ANALYTICS GROUP	1
Data Scientist, Senior	1
Sr. Data Scientist, Cyber-Security LT Contract	1
MongoDB Data Engineer II	1
Data Scientist - Bioinformatics	1
Principal Machine Learning Scientist	1
Data Analyst / Scientist	1
Data Scientist - Research	1
R&D Data Analysis Scientist	1
Analytics Consultant	1
Director, Data Science	1
R&D Sr Data Scientist	1
Customer Data Scientist/Sales Engineer (Bay	1
Jr. Business Data Analyst	1
Data Management Specialist	1
E-Commerce Data Analyst	1
Data Engineer I - Azure	1
Insurance Data Scientist	1
Data Modeler	1
Data Scientist, Rice University	1
Senior Research Scientist - Embedded System Development for DevOps	1
Financial Data Analyst	1
Ag Data Scientist	1
Data Scientist II	1
Project Scientist	1
Data Analytics Manager	1
Senior Machine Learning (ML) Engineer / Data Scientist - Cyber Security Analytics	1
Associate Scientist	1
Scientist 2, QC Viral Vector	1

```
Data Scientist/ML Engineer                    1
Sr. Data Scientist                            1
Data Engineer 4 - Contract                    1
Data Analyst - Asset Management               1
Machine Learning Engineer (NLP)               1
Name: Job Title, dtype: int64
```

In [18]:

```
df1['Location'].value_counts()
```

Out[18]:

```
New York, NY                  55
San Francisco, CA             49
Cambridge, MA                 47
Chicago, IL                   32
Boston, MA                    23
San Jose, CA                  13
Pittsburgh, PA                12
Washington, DC                11
Rockville, MD                 11
Winston-Salem, NC             10
Richland, WA                  10
Herndon, VA                   10
Indianapolis, IN               9
San Diego, CA                  9
Mountain View, CA              8
Austin, TX                     8
South San Francisco, CA        8
Rochester, NY                  7
Palo Alto, CA                  7
Salt Lake City, UT             6
Huntsville, AL                 6
Marlborough, MA                6
Phoenix, AZ                    6
Charlotte, NC                  6
Chantilly, VA                  6
Dallas, TX                     6
Gaithersburg, MD               6
Philadelphia, PA               5
Worcester, MA                  5
Milwaukee, WI                  5
Denver, CO                     5
Nashville, TN                  5
Springfield, MA                5
Cincinnati, OH                 5
Seattle, WA                    5
Durham, NC                     4
Houston, TX                    4
Omaha, NE                      4
Fort Belvoir, VA               4
Owensboro, KY                  4
```

Arlington, VA	4
Millville, DE	4
Los Angeles, CA	4
Redlands, CA	4
Clearwater, FL	4
Burbank, CA	4
Highland, CA	4
Burleson, TX	4
Scotts Valley, CA	4
Hoopeston, IL	4
Knoxville, TN	4
Orange, CA	3
Riverton, UT	3
San Antonio, TX	3
Lenexa, KS	3
Hartford, CT	3
Silver Spring, MD	3
Tampa, FL	3
Annapolis Junction, MD	3
McLean, VA	3
Vail, CO	3
Concord, CA	3
Albuquerque, NM	3
Natick, MA	3
Atlanta, GA	3
Baltimore, MD	3
Springfield, VA	3
Fort Lee, NJ	3
Richfield, OH	3
Armonk, NY	3
Beavercreek, OH	3
Ithaca, NY	3
Hampton, VA	3
Quincy, MA	3
Marietta, GA	3
Clovis, CA	2
Newton, MA	2
Alameda, CA	2
Lafayette, LA	2
Allentown, PA	2
Plymouth Meeting, PA	2
Franklin, TN	2
Dublin, CA	2
Aliso Viejo, CA	2
Bedford, MA	2
Santa Clara, CA	2
Jersey City, NJ	2
Blue Bell, PA	2
Lewes, DE	2
Cedar Rapids, IA	2

Phila, PA	2
Des Moines, IA	2
West Palm Beach, FL	2
Westlake, OH	2
Exton, PA	2
Green Bay, WI	2
Alexandria, VA	2
Columbia, MO	2
Orlando, FL	2
Chandler, AZ	2
Springfield, MO	2
Hamilton, NJ	2
Woodbridge, NJ	2
Bellevue, WA	2
Vancouver, WA	2
Detroit, MI	2
Sunnyvale, CA	2
Ipswich, MA	2
Minneapolis, MN	2
Fremont, CA	2
New Orleans, LA	2
Louisville, KY	2
Peoria, IL	2
Fort Lauderdale, FL	2
Saint Louis, MO	2
Coraopolis, PA	2
Charlottesville, VA	2
Maryland Heights, MO	2
Hillsboro, OR	2
Groton, CT	2
Cupertino, CA	2
San Rafael, CA	2
Winter Park, FL	2
Dearborn, MI	2
Madison, WI	2
Plano, TX	2
West Reading, PA	2
Frederick, MD	2
Ewing, NJ	2
Ann Arbor, MI	1
Carle Place, NY	1
Santa Barbara, CA	1
Emeryville, CA	1
Bloomington, IL	1
Long Beach, NY	1
Allendale, NJ	1
Longmont, CO	1
King of Prussia, PA	1
Irvine, CA	1
Woburn, MA	1

Scottsdale, AZ	1
Glen Burnie, MD	1
Reston, VA	1
Birmingham, AL	1
Logan, UT	1
Newark, NJ	1
Roanoke, VA	1
Arvada, CO	1
Milpitas, CA	1
Brisbane, CA	1
Watertown, MA	1
Cambridge, MD	1
Corvallis, OR	1
Waltham, MA	1
Holyoke, MA	1
Foster City, CA	1
Framingham, MA	1
Tacoma, WA	1
Lake Forest, IL	1
Valencia, CA	1
Boise, ID	1
Oakland, CA	1
Ashburn, VA	1
Raleigh, NC	1
Providence, RI	1
Fort Worth, TX	1
Port Washington, NY	1
Portland, OR	1
Cherry Hill, NJ	1
Harrisburg, PA	1
Meridian, ID	1
Parlier, CA	1
Dayton, OH	1
San Mateo, CA	1
Sheboygan, WI	1
Novato, CA	1
Aurora, CO	1
Chattanooga, TN	1
Oak Ridge, TN	1
Agoura Hills, CA	1
Pella, IA	1
San Ramon, CA	1
Laurel, MD	1
Linthicum, MD	1
Miami, FL	1
Landover, MD	1
Patuxent River, MD	1
Suitland, MD	1
Syracuse, NY	1
Southfield, MI	1

```
Matawan, NJ                          1
Lyndhurst, NJ                        1
Atlanta, IN                          1
Alabaster, AL                        1
Portsmouth, VA                       1
Santa Fe Springs, Los Angeles, CA    1
Kansas City, MO                      1
Columbia, SC                         1
Red Bank, NJ                         1
Olympia, WA                          1
Name: Location, dtype: int64
```

In [19]:

```
df1['Size'].value_counts()
```

Out[19]:

```
1001 - 5000      150
501 - 1000       134
10000+           130
201 - 500        117
51 - 200          94
5001 - 10000      76
1 - 50            31
unknown           10
Name: Size, dtype: int64
```

In [20]:

```
df1['Founded'].value_counts()
```

Out[20]:

```
-1       50
 2010    32
 2008    31
 1996    27
 2006    24
 2012    21
 2011    19
 1958    18
 2007    18
 1984    18
 2002    18
 2015    16
 2013    15
 1875    14
 1997    14
 1851    14
 1781    14
 2014    13
 1965    12
 2017    12
 1999    12
 2005    10
```

1912 10
2003 10
2000 10
1935 10
1961 9
1913 9
1982 9
1981 9
1977 8
1995 8
1939 8
1989 8
1969 8
1968 8
1976 8
1849 7
1988 7
1992 7
1948 6
2004 6
1986 6
1993 6
2009 6
1870 6
1967 5
1966 5
2016 5
1973 5
1852 5
1964 4
1830 4
1991 4
1994 4
1925 4
1915 4
1947 3
1970 3
1943 3
1922 3
1972 3
2001 3
1978 3
1863 3
1885 3
1937 3
1990 3
1998 3
1987 2
1974 2
1952 2

```
1856    2
1983    2
1962    2
1980    2
1954    2
1975    2
1951    2
2019    2
1846    2
1928    2
1914    1
1812    1
1985    1
1899    1
1979    1
1929    1
1927    1
1945    1
1744    1
1902    1
1850    1
1887    1
1883    1
1917    1
1930    1
1860    1
1942    1
1878    1
1971    1
1889    1
Name: Founded, dtype: int64
```

In [21]:

```
df1['Industry'].value_counts()
```

Out[21]:

```
Biotech & Pharmaceuticals                  112
Insurance Carriers                          63
Computer Hardware & Software                59
IT Services                                 50
Health Care Services & Hospitals            49
Enterprise Software & Network Solutions     42
Internet                                    29
Consulting                                  29
Aerospace & Defense                         25
Advertising & Marketing                     25
Consumer Products Manufacturing             20
Research & Development                      19
Colleges & Universities                     16
Energy                                      14
Banks & Credit Unions                       12
```

```
Federal Agencies                          11
-1                                        10
Staffing & Outsourcing                    10
Travel Agencies                            8
Lending                                    8
Food & Beverage Manufacturing              8
Financial Analytics & Research             8
Real Estate                                8
Security Services                          7
Insurance Agencies & Brokerages            6
Religious Organizations                    6
Department, Clothing, & Shoe Stores        6
Investment Banking & Asset Management      5
Architectural & Engineering Services       4
K-12 Education                             4
Gas Stations                               4
Gambling                                   4
Industrial Manufacturing                   4
Telecommunications Services                4
Logistics & Supply Chain                   4
Financial Transaction Processing           4
Wholesale                                  3
Stock Exchanges                            3
Social Assistance                          3
Construction                               3
Transportation Management                  3
Video Games                                3
Education Training Services                3
Consumer Product Rental                    3
Telecommunications Manufacturing           2
Sporting Goods Stores                      2
TV Broadcast & Cable Networks              2
Brokerage Services                         2
Metals Brokers                             2
Accounting                                 1
Health Care Products Manufacturing         1
Health, Beauty, & Fitness                  1
Farm Support Services                      1
Auctions & Galleries                       1
Trucking                                   1
Mining                                     1
Other Retail Stores                        1
Motion Picture Production & Distribution   1
Transportation Equipment Manufacturing     1
Beauty & Personal Accessories Stores       1
Name: Industry, dtype: int64
```

In [24]:

```
df1['Lower Salary'].value_counts()
```

Out[24]:

43	22
65	20
61	18
80	18
52	18
49	18
81	17
74	16
63	16
56	16
86	15
60	15
54	14
42	14
71	13
44	12
100	11
37	11
68	11
110	11
64	10
75	10
50	10
39	10
83	10
76	10
55	9
59	9
108	9
102	9
72	9
85	9
31	8
90	8
40	8
82	8
35	8
48	8
97	7
120	7
69	7
62	7
150	7
107	7
67	6
95	6
91	6
116	6
53	6
32	6

84	6
79	6
45	6
38	5
113	5
92	5
111	5
87	5
66	5
34	5
77	5
105	5
109	5
124	4
58	4
114	4
57	4
94	4
47	4
36	4
93	4
118	4
73	4
190	3
101	3
138	3
127	3
202	3
78	3
70	3
33	3
117	3
41	3
121	3
200	3
89	3
99	3
139	3
20	3
98	2
126	2
27	2
119	2
135	2
106	2
125	2
115	2
158	2
88	2
132	2

```
96      2
130     2
112     2
51      2
131     2
46      1
15      1
29      1
176     1
26      1
129     1
136     1
171     1
Name: Lower Salary, dtype: int64
```

In [25]:

```
df1['Upper Salary'].value_counts()
```

Out[25]:

```
140    16
119    15
110    15
124    15
113    13
127    13
86     12
173    12
101    12
139    11
85     11
142    10
62     10
97     10
134    10
160    10
123     9
99      9
133     9
112     9
129     8
105     8
149     8
143     8
132     8
126     7
148     7
115     7
78      7
95      7
82      7
96      7
```

91	7
81	7
172	7
135	7
93	7
144	7
70	6
76	6
100	6
71	6
92	6
68	6
182	6
66	6
106	6
80	6
111	6
137	6
211	6
179	6
158	5
159	5
125	5
130	5
98	5
167	5
157	5
220	5
72	5
109	5
146	5
52	5
114	5
120	5
136	4
121	4
90	4
189	4
208	4
116	4
89	4
102	4
150	4
117	4
175	4
147	4
180	4
176	4
171	4
178	4

161	4
166	4
153	4
224	3
108	3
59	3
155	3
58	3
64	3
199	3
306	3
196	3
88	3
200	3
57	3
194	3
77	3
55	3
154	3
141	3
118	3
206	3
204	3
190	3
145	3
65	3
203	2
210	2
138	2
221	2
61	2
151	2
183	2
177	2
60	2
239	2
181	2
228	2
238	2
187	2
69	2
48	2
35	2
122	2
207	2
250	2
49	2
185	2
198	2
162	2

```
202    2
73     2
174    2
87     1
79     1
188    1
272    1
165    1
104    1
84     1
215    1
201    1
184    1
107    1
289    1
231    1
103    1
16     1
223    1
50     1
63     1
163    1
74     1
193    1
209    1
275    1
39     1
222    1
164    1
67     1
Name: Upper Salary, dtype: int64
```

In [26]:

```
df1['Avg Salary(K)'].value_counts()
```

Out[26]:

```
87.5     12
140.0    11
81.0     11
85.0     10
107.5    10
56.5     10
84.5     10
107.0    10
87.0      9
120.0     9
154.5     8
109.0     8
70.5      8
76.5      8
100.0     7
```

65.0	7
85.5	7
95.0	7
121.0	7
62.5	7
61.0	7
114.5	7
77.5	7
80.5	7
54.0	6
51.5	6
139.5	6
68.5	6
106.5	6
124.0	6
52.5	6
112.5	6
96.0	6
61.5	6
94.5	6
98.0	5
75.5	5
66.5	5
128.5	5
44.5	5
99.0	5
93.5	5
114.0	5
99.5	5
92.0	5
111.5	5
80.0	5
98.5	5
101.0	5
73.0	5
113.5	5
73.5	5
103.5	5
65.5	5
139.0	5
55.0	4
91.5	4
124.5	4
161.5	4
64.0	4
110.5	4
71.5	4
162.0	4
72.5	4
100.5	4

```
138.5     4
147.0     4
117.5     4
48.5      4
86.5      4
109.5     4
130.0     4
97.5      4
84.0      4
105.5     4
142.5     4
115.0     4
90.0      4
69.5      4
173.0     3
155.0     3
142.0     3
254.0     3
169.0     3
137.0     3
45.5      3
123.5     3
50.0      3
110.0     3
122.5     3
47.0      3
59.0      3
140.5     3
125.0     3
134.5     3
181.0     3
205.0     3
60.0      3
104.5     3
90.5      3
116.5     3
102.5     3
49.0      3
53.5      3
150.5     3
128.0     3
132.5     3
74.0      3
96.5      3
133.0     3
120.5     3
122.0     3
153.0     3
62.0      3
67.0      3
```

167.5	3
143.5	3
70.0	3
164.5	2
63.0	2
82.0	2
71.0	2
149.5	2
44.0	2
72.0	2
64.5	2
129.5	2
103.0	2
83.0	2
127.5	2
37.5	2
113.0	2
194.0	2
172.0	2
86.0	2
66.0	2
94.0	2
74.5	2
136.5	2
194.5	2
27.5	2
119.0	2
43.0	2
115.5	2
51.0	2
180.0	2
60.5	2
151.5	2
97.0	2
145.0	2
184.5	2
95.5	2
92.5	2
88.0	2
147.5	2
106.0	2
81.5	2
168.0	2
63.5	2
47.5	2
143.0	2
164.0	2
171.5	2
42.0	2
225.0	2

```
148.0    2
146.5    2
89.0     2
153.5    2
102.0    2
48.0     2
112.0    2
76.0     2
91.0     2
79.5     2
108.0    2
59.5     1
165.0    1
68.0     1
145.5    1
58.0     1
29.5     1
82.5     1
177.0    1
133.5    1
121.5    1
118.5    1
58.5     1
146.0    1
221.5    1
53.0     1
163.5    1
101.5    1
237.5    1
127.0    1
137.5    1
77.0     1
93.0     1
162.5    1
163.0    1
83.5     1
15.5     1
88.5     1
69.0     1
79.0     1
41.5     1
105.0    1
118.0    1
89.5     1
40.5     1
39.5     1
134.0    1
119.5    1
78.0     1
157.0    1
```

```
174.0     1
179.5     1
232.5     1
111.0     1
Name: Avg Salary(K), dtype: int64
```

I had some trouble loading this cleaned data into Tableau! Surprisingly Tableau wasn't recognizing the columns, it was grabbing random sections of text. Hmm.

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

V. Conclusion

This assignment was a bit difficult! I found it difficult managing the combination of both a brand new software and a dataset I had to pick and clean myself. The combination made the assignment quite difficult to get perfectly correct. I will continue working with Tableau though! I'm sure it is just the beginning of my Tableau career.

Thank you! Jeremy

VI. References

1) From the Experts PDF, Week 8

2) Tableau Tutorial. (2020). Tutorials Point. Retrieved May 1, 2022, from https://www.tutorialspoint.com/tableau/

3) voter registration dataset, https://github.com/fivethirtyeight/data/tree/master/voter-registration

4) data scientist salaries dataset, https://www.kaggle.com/datasets/nikhilbhathi/data-scientist-salary-us-glassdoor

In [ ]: