Employee Survey Analysis

Essential Questions

- Where should we spend our resources to make the biggest impact in lowering employee attrition?
- Which companies are most at risk?
- Which companies are doing well and what can we do to improve that further?

Understanding the Data

FIELD	FORMAT	DESCRIPTION
EXECUTIVE	Character	The executive vice president that employee reports up to
COMPANY	Character	The business that the employee works for
GROUP_LAYER	Character	The layer within the organization structure that the employee resides (Ex. CEO = layer 1; VP = layer 2…)
CULTURE	Integer	My workgroup has a culture in which employees appreciate the differences that people bring to the workplace.
CAREER	Integer	I have good career opportunities at this company.
COMMUNICATION	Integer	This company does a good job of communicating with employees.
GROWTH	Integer	I have good opportunities to learn and grow at this company.
MANAGER	Integer	I would recommend my manager/supervisor to others.
RESPECT	Integer	I am treated with respect and dignity.
TEAM	Integer	I am satisfied with my work team.
SATISFACTION	Integer	How happy are you working here?

Looking at the Data

In [246]:
```
import numpy as np
import pandas as pd
import seaborn as sns
```

In [247]:
```
df = pd.read_csv("Employee_Survey_Raw_Data.csv")
```

In [248]:
```
df.head()
```

Out[248]:

	Executive	Company	Group_Layer	Culture	Career	Communication	Growth	Manager	Respect	Team	Satisfaction
0	Achim Plass	Pantheon	`03-04	4	3	4	3	3	4	4	3
1	Achim Plass	Pantheon	`03-04	2	3	2	4	3	3	4	3
2	Achim Plass	Pantheon	`03-04	3	5	5	5	5	5	5	5
3	Andreas Hargrove	Plexus	`03-04	4	3	5	4	5	5	3	2
4	Andreas Hargrove	Plexus	`03-04	4	4	4	4	2	4	4	4

In [249]:
```
df.shape
```

Out[249]:
```
(174, 11)
```

In [250]:
```
df.Company.value_counts()
```

Out[250]:
```
Sparta        39
BrightStar    36
Pantheon      33
Plexus        30
Cyberspace    24
Fluent        12
```

Summary Statistics

```
In [11]: df.describe()
```

Out[11]:

	Culture	Career	Communication	Growth	Manager	Respect	Team	Satisfaction
count	174.000000	174.000000	174.000000	174.000000	174.000000	174.000000	174.000000	174.000000
mean	4.022989	3.741379	3.810345	3.959770	4.178161	4.229885	4.137931	3.925287
std	0.973368	1.110543	1.016458	0.945688	1.079170	0.927439	0.914525	0.979645
min	-1.000000	-1.000000	-1.000000	-1.000000	-1.000000	-1.000000	-1.000000	-1.000000
25%	4.000000	3.000000	3.000000	3.250000	4.000000	4.000000	4.000000	3.000000
50%	4.000000	4.000000	4.000000	4.000000	4.000000	4.000000	4.000000	4.000000
75%	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000
max	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000

Data Cleaning

Missing Value Treatment

In [252]:
```
# Replace -1 to Nan
df = df.replace([-1], np.nan)
```

In [253]:
```
# Replace Nan to the median value of the feature column
df = df.fillna(df.median())
```

In [254]:
```
df.isnull().sum()
```

Out[254]:
```
Executive         0
Company           0
Group_Layer       0
Culture           0
Career            0
Communication     0
Growth            0
Manager           0
Respect           0
Team              0
Satisfaction      0
dtype: int64
```

Data Cleaning

```
In [255]: df.describe()
```

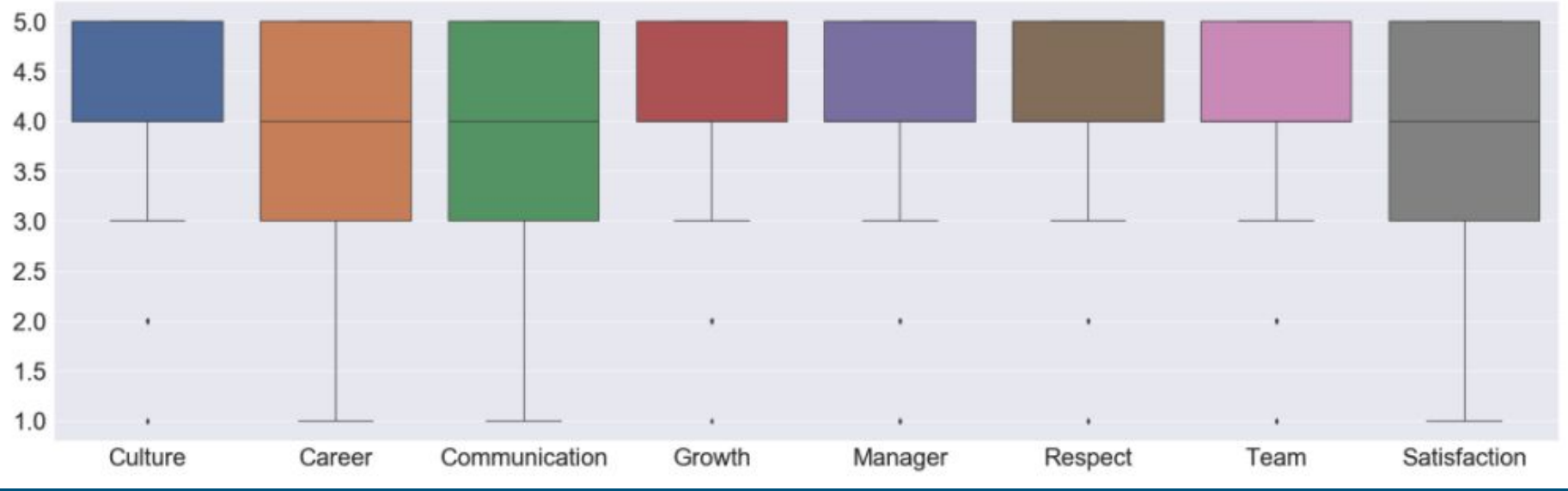
Out[255]:

	Culture	Career	Communication	Growth	Manager	Respect	Team	Satisfaction
count	174.000000	174.000000	174.000000	174.000000	174.000000	174.000000	174.000000	174.000000
mean	4.051724	3.770115	3.839080	3.988506	4.235632	4.258621	4.166667	3.954023
std	0.894862	1.050198	0.948056	0.866783	0.922860	0.837565	0.826465	0.904811
min	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
25%	4.000000	3.000000	3.000000	4.000000	4.000000	4.000000	4.000000	3.000000
50%	4.000000	4.000000	4.000000	4.000000	4.000000	4.000000	4.000000	4.000000
75%	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000
max	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000

Distribution and Outliers

```
sns.set(font_scale=3)
sns.catplot(data=df,kind='box',height=10,aspect=3)
```

```
<seaborn.axisgrid.FacetGrid at 0x1a274cf950>
```



Relationship Analysis - Correlation Table

Relationship Analysis

In [272]:
```
corr = df.corr()
corr
```

Out[272]:

	Culture	Career	Communication	Growth	Manager	Respect	Team	Satisfaction
Culture	1.000000	0.338714	0.418672	0.336122	0.356126	0.467919	0.496303	0.495548
Career	0.338714	1.000000	0.392246	0.765428	0.342491	0.514843	0.390705	0.518042
Communication	0.418672	0.392246	1.000000	0.356476	0.294643	0.409409	0.270500	0.395634
Growth	0.336122	0.765428	0.356476	1.000000	0.364714	0.489804	0.381932	0.478392
Manager	0.356126	0.342491	0.294643	0.364714	1.000000	0.541399	0.440827	0.463009
Respect	0.467919	0.514843	0.409409	0.489804	0.541399	1.000000	0.546956	0.595465
Team	0.496303	0.390705	0.270500	0.381932	0.440827	0.546956	1.000000	0.566856
Satisfaction	0.495548	0.518042	0.395634	0.478392	0.463009	0.595465	0.566856	1.000000

Relationship Analysis - Correlation Table

	Culture	Career	Communication	Growth	Manager	Respect	Team	Satisfaction
Culture	1.000000	0.338714	0.418672	0.336122	0.356126	0.467919	0.496303	0.495548
Career	0.338714	1.000000	0.392246	0.765428	0.342491	0.514843	0.390705	0.518042
Communication	0.418672	0.392246	1.000000	0.356476	0.294643	0.409409	0.270500	0.395634
Growth	0.336122	0.765428	0.356476	1.000000	0.364714	0.489804	0.381932	0.478392
Manager	0.356126	0.342491	0.294643	0.364714	1.000000	0.541399	0.440827	0.463009
Respect	0.467919	0.514843	0.409409	0.489804	0.541399	1.000000	0.546956	0.595465
Team	0.496303	0.390705	0.270500	0.381932	0.440827	0.546956	1.000000	0.566856
Satisfaction	0.495548	0.518042	0.395634	0.478392	0.463009	0.595465	0.566856	1.000000

- Values from -1:1 indicating the measure of strength of the relationship between two variables (1 = strongest positive relationship, -1 = strongest negative relationship)
- Satisfaction is the keystone metric
- Variables with strong correlation with Satisfaction can show which metrics impact the satisfaction score the most
- Respect and Team have the strongest relationship with satisfaction while Communication has the weakest

Relationship Analysis - Heat Map

```
sns.heatmap(corr)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x1a334974d0>
```

Multiple Linear Regression

```
: from sklearn import linear_model

X = df[['Culture','Career','Communication','Growth','Manager','Respect','Team']]
y = df['Satisfaction']

regr = linear_model.LinearRegression()
regr.fit(X, y)

print("Coefficients: " ,regr.coef_)
print("Intercept: ", regr.intercept_)
print("R Squared Value: ", regr.score(X,y))
```

```
Coefficients:  [0.14393909 0.15266178 0.07166857 0.04401203 0.09846918 0.2195831
 0.25739613]
Intercept:  -0.0800988377905143
R Squared Value:  0.5130276063085571
```

Multiple Linear Regression Model

- Satisfaction = .1439 Culture + .1527 Career + .0717 Communication + .0440 Growth + .0985 Manager + .2196 Respect + .2574 Team - .08

- 51.3% of the variability in the model is explained by the independent variables
  - We are simply looking at the relationship between the variables and not predicting so this is less concerning

Where should we spend our resources?

- People care about Respect and Team the most
  - Each point increase in Respect and Team led to an increase of .2196 and .2574 points in Satisfaction

- Communication and Growth are the least important factors for job satisfaction
  - Each point increase in Communication and Growth led to only an increase of .0717 and .0440 points in Satisfaction

- This matches our intuition from the heatmap as these factors had the biggest and smallest correlation coefficients

Strengths and Weaknesses of the Companies

```
df.pivot_table(index='Company', aggfunc='mean')
```

Company	Career	Communication	Culture	Growth	Manager	Respect	Satisfaction	Team
BrightStar	3.416667	3.527778	4.000000	3.777778	3.916667	4.055556	3.777778	3.916667
Cyberspace	3.791667	4.208333	3.916667	3.916667	4.291667	4.250000	4.125000	4.083333
Fluent	3.750000	3.416667	4.333333	4.166667	4.416667	4.583333	4.166667	4.333333
Pantheon	3.757576	3.939394	4.181818	3.909091	4.151515	4.393939	4.121212	4.484848
Plexus	3.733333	3.800000	3.933333	3.966667	4.333333	4.066667	3.600000	4.000000
Sparta	4.128205	3.974359	4.076923	4.256410	4.435897	4.384615	4.076923	4.256410

	Culture	Career	Communication	Growth	Manager	Respect	Team	Satisfaction
count	174.000000	174.000000	174.000000	174.000000	174.000000	174.000000	174.000000	174.000000
mean	4.051724	3.770115	3.839080	3.988506	4.235632	4.258621	4.166667	3.954023

Which companies are most at risk?

```
df.pivot_table(index='Company', aggfunc='mean')
```

Company	Career	Communication	Culture	Growth	Manager	Respect	Satisfaction	Team
BrightStar	3.416667	3.527778	4.000000	3.777778	3.916667	4.055556	3.777778	3.916667
Cyberspace	3.791667	4.208333	3.916667	3.916667	4.291667	4.250000	4.125000	4.083333
Fluent	3.750000	3.416667	4.333333	4.166667	4.416667	4.583333	4.166667	4.333333
Pantheon	3.757576	3.939394	4.181818	3.909091	4.151515	4.393939	4.121212	4.484848
Plexus	3.733333	3.800000	3.933333	3.966667	4.333333	4.066667	3.600000	4.000000
Sparta	4.128205	3.974359	4.076923	4.256410	4.435897	4.384615	4.076923	4.256410

	Culture	Career	Communication	Growth	Manager	Respect	Team	Satisfaction
count	174.000000	174.000000	174.000000	174.000000	174.000000	174.000000	174.000000	174.000000
mean	4.051724	3.770115	3.839080	3.988506	4.235632	4.258621	4.166667	3.954023

- BrightStar and Plexus have satisfaction scores of 3.78 and 3.60 respectively which are lower than the overall average of 3.95
- Respect and Team scores (BrightStar = 4.06, 3.92) (Plexus = 4.07, 4) are lower than overall average of 4.26, 4.17 respectively
- Low scores on the biggest impact metrics are a concern and must be addressed

Which companies are doing well?

```
df.pivot_table(index='Company', aggfunc='mean')
```

Company	Career	Communication	Culture	Growth	Manager	Respect	Satisfaction	Team
BrightStar	3.416667	3.527778	4.000000	3.777778	3.916667	4.055556	3.777778	3.916667
Cyberspace	3.791667	4.208333	3.916667	3.916667	4.291667	4.250000	4.125000	4.083333
Fluent	3.750000	3.416667	4.333333	4.166667	4.416667	4.583333	4.166667	4.333333
Pantheon	3.757576	3.939394	4.181818	3.909091	4.151515	4.393939	4.121212	4.484848
Plexus	3.733333	3.800000	3.933333	3.966667	4.333333	4.066667	3.600000	4.000000
Sparta	4.128205	3.974359	4.076923	4.256410	4.435897	4.384615	4.076923	4.256410

	Culture	Career	Communication	Growth	Manager	Respect	Team	Satisfaction
count	174.000000	174.000000	174.000000	174.000000	174.000000	174.000000	174.000000	174.000000
mean	4.051724	3.770115	3.839080	3.988506	4.235632	4.258621	4.166667	3.954023

- Pantheon, Fluent, Cyberspace have the strongest satisfaction scores
- Pantheon's team score was strong but scored below average in Respect
- Cyberspace's Respect and Team scores were average and below average respectively. Boosting these metrics would drastically improve satisfaction

Potential Areas of Error

- Outliers could impact data
- Sample size of certain companies were small like Fluent which only had 12 data points
- Missing data

Further Interest

- If we had attrition data in the form “yes” or “no”
  - Which companies had the highest employee attrition?
  - Does employee rank affect employee attrition?
  - Attrition rate: # of employees that left / Avg # of employees

- Is there a statistically significant difference in attrition between men and women?

Thank you