TEAM 3

Employee Satisfaction Analysis

GLASSDOOR & YAHOO FINANCE

Meet the Anatysts

TEAM 3

Hannah Wilt

John Mungenast

Matt Peckman

AGENDA

01 The Problem
02 Previous Efforts
03 Introduction of Datasets
04 Cleaning & Transformation
05 Data Analysis
06 Challenges
07 Potential Expansion for Problem

The Business Problem

Should companies invest more into caring about their employees and things that are important to them because it positively impacts their financial performance and stock price?

PREVIOUS EFFORTS MADE

Newcastle University Study, 2017

- Analyzed work-life balance and stress at work
- Found 94% felt less stressed with W/L balance
- Reported 79% increase in engagement
- Placed emphasis on the value add to the firm

TAKING IT A STEP FURTHER

How Team 3 Builds on Previous Study

Time-series analysis to look at overall rating overtime

Look at financial performance on the stock market

TAKING IT A STEP FURTHER

How Team 3 Builds on Previous Study

Determining accuracy with firms rated +/-

Determining if companies investing into employees satisfaction increases likelihood of stock increase

INTRODUCING THE *Datasets*

glassdoor kaggle

STRENGTHS	WEAKNESSES
Open source	Needed cleaned
Easy to download and manipulate	Needed to remove all private companies (not publicly traded)
Structured data for DT and RF	Contained unstructured data

yahoo! finance

STRENGTHS	WEAKNESSES
Provided financial data	Needed cleaned
Data could easily be manipulated to display trends	Had to append ticker data Into kaggle dataset
Open source	Large file (slow processing)

METHODOLOGY & STUDY DESIGN

PRESCRIPTIVE

TO PROVIDE DECISION MAKING RECOMMENDATIONS

METHOD: DECISION TREE

PREDICTIVE

TO FORECAST FUTURE TRENDS

METHOD: TIME SERIES ANALYSIS (ML)

DIAGNOSTIC

TO DETERMINE CAUSE OF EVENTS BY IDENTIFYING PATTERNS

METHOD: REGRESSION ANALYSIS

WHAT DO WE PLAN TO LEARN?

PRESCRIPTIVE

METHOD: DECISION TREE

TO SEE WHICH VARIABLES WERE MORE LIKELY TO LEAD TO POSITIVE OVERALL RATING

PREDICTIVE

METHOD: TIME SERIES ANALYSIS (ML)

TO SEE COMPANIES OVERALL RATING OVER TIME AND DETERMINE ANY FUTURE TRENDS

DIAGNOSTIC

METHOD: REGRESSION ANALYSIS

TO SEE IF THERE IS ANY CORRELATION BETWEEN VARIABLES AND COMPANY MARKET PRICE

CLEANING THE kaggle DATASET

- Fuzzy joined to list of tickers
- Appended NYSE Tickers
- Removed NaN values
- Formatted Dates

	date_review	firm	Symbol	work_life_balance	overall_rating
0	2020-09-16	McDonald-s	MCD	3.0	2
1	2020-09-19	McDonald-s	MCD	5.0	4
2	2020-09-22	McDonald-s	MCD	5.0	3
3	2020-09-30	McDonald-s	MCD	5.0	3
4	2020-10-05	McDonald-s	MCD	5.0	3
5	2020-10-08	McDonald-s	MCD	4.0	4
6	2020-10-09	McDonald-s	MCD	3.0	2
7	2020-10-11	McDonald-s	MCD	4.0	4
8	2020-10-18	McDonald-s	MCD	5.0	3
9	2020-10-19	McDonald-s	MCD	2.0	3

PREPPING THE DATASET WITH POWER BI

A1 | fx | firm

	A	B	C	D	location
1	firm	date_review	job_title	current	
2	Oracle	41389	Anonymous	Current Employee	
3	Oracle	41389	Anonymous	Current Employee	
4	Oracle	41403	Anonymous	Current Employee	
5	Oracle	41408	Anonymous	Current Employee	
6	Oracle	41408	Anonymous	Current Employee	
7	Oracle	41424	Anonymous	Current Employee	
8	Oracle	41669	Anonymous	Current Employee	
9	Oracle	41883	Anonymous	Current Employee	
10	Oracle	42115	Anonymous	Current Employee	
11	Oracle	42121	Anonymous	Current Employee	
12	Oracle	42173	Anonymous	Current Employee	
13	Oracle	42184	Anonymous	Current Employee	
14	Oracle	42184	Anonymous	Current Employee	
15	Oracle	42206	Anonymous	Current Employee	
16	Oracle	42206	Anonymous	Current Employee	
17	Oracle	42229	Anonymous	Current Employee	
18	Oracle	42233	Anonymous	Current Employee	
19	Oracle	42249	Anonymous	Current Employee	
20	Oracle	42263	Anonymous	Current Employee	
21	Oracle	42296	Anonymous	Current Employee	
22	Oracle	42296	Anonymous	Current Employee	
23	Oracle	42321	Anonymous	Current Employee	
24	Oracle	42333	Anonymous	Current Employee	
25	Oracle	42380	Anonymous	Current Employee	
26	Oracle	42380	Anonymous	Current Employee	
27	Oracle	42389	Anonymous	Current Employee	
28	Oracle	42391	Anonymous	Current Employee	
29	Oracle	42392	Anonymous	Current Employee	
30	Oracle	42394	Anonymous	Current Employee	
31	Oracle	42451	Anonymous	Current Employee	
32	Oracle	42451	Anonymous	Current Employee	
33	Oracle	42452	Anonymous	Current Employee	

FUZZY JOIN WILL PROVIDE US WITH 80% MATCHING IN THE NAME IN POWER BI

B1 | fx | Name

	A	B
1	Symbol	Name
2	A	Agilent Technologies Inc. Common Stock
3	AA	Alcoa Corporation Common Stock
4	AACG	ATA Creativity Global American Depositary Shares
5	AACI	Armada Acquisition Corp. I Common Stock
6	AACIW	Armada Acquisition Corp. I Warrant
7	AACT	Ares Acquisition Corporation II Class A Ordinary Shares
8	AADI	Aadi Bioscience Inc. Common Stock
9	AAIC	Arlington Asset Investment Corp Class A (new)
10	AAIC^B	Arlington Asset Investment Corp 7.00%
11	AAIC^C	Arlington Asset Investment Corp 8.250% Seies C Fixed-to-Floating Rate Cumulative Redeemable Preferred Stock
12	AAIN	Arlington Asset Investment Corp 6.000% Senior Notes Due 2026
13	AAL	American Airlines Group Inc. Common Stock
14	AAMC	Altisource Asset Management Corp Com
15	AAME	Atlantic American Corporation Common Stock
16	AAN	Aarons Holdings Company Inc. Common Stock
17	AAOI	Applied Optoelectronics Inc. Common Stock
18	AAON	AAON Inc. Common Stock
19	AAP	Advance Auto Parts Inc.
20	AAPL	Apple Inc. Common Stock
21	AAT	American Assets Trust Inc. Common Stock
22	AAU	Almaden Minerals Ltd. Common Shares
23	AB	AllianceBernstein Holding L.P. Units
24	ABAT	American Battery Technology Company Common Stock
25	ABBV	AbbVie Inc. Common Stock
26	ABCB	Ameris Bancorp Common Stock
27	ABCL	AbCellera Biologics Inc. Common Shares
28	ABCM	Abcam plc American Depositary Shares
29	ABEO	Abeona Therapeutics Inc. Common Stock
30	ABEV	Ambev S.A. American Depositary Shares (Each representing 1 Common Share)

WE ONLY WANT THE TICKERS THAT MATCH THAT 80% THRESHOLD, SO WE WILL USE AN INNER JOIN TO RETURN ONLY MATCHING VALUES FROM BOTH TABLES

RETRIEVING THE REGULAR MARKET PRICE

```
tickers = kaggle['Symbol'].unique()

print(tickers)
```

```
['MCD' 'ORAN' 'ORCL' 'MS' 'TRI' 'CRM' 'AAPL' 'ARMK' 'MSFT' 'AXP' 'HSBC'
 'YEXT' 'LYG' 'WTW' 'DVA' 'PRSO' 'KFY' 'SPWR' 'WIT' 'CSCO' 'IBIO' 'DT'
 'AWRE']
```

```
import urllib.request
import json
from pandas import json_normalize

raw_finance = []

beg_url = "https://query1.finance.yahoo.com/v8/finance/chart/"
end_url = "?metrics=high?&interval=3mo&range=5y"
```

RETURN URLS WITH EACH TICKER AND FINANCIAL INFORMATION

```
for ticker in tickers:
    full_url = beg_url + ticker + end_url

    print(full_url)

    result = urllib.request.urlopen(full_url).read()
    result_dict = json.loads(result)
    fin_data = json_normalize(result_dict['chart']['result'])

    try:
        market_price = float(fin_data['meta.regularMarketPrice'][0])
        raw_finance.append(market_price)
    except (KeyError, ValueError):
        raw_finance.append(None)

print(raw_finance)
```

```
https://query1.finance.yahoo.com/v8/finance/chart/MCD?metrics=high?&interval=3mo&range=5y
https://query1.finance.yahoo.com/v8/finance/chart/ORAN?metrics=high?&interval=3mo&range=5y
https://query1.finance.yahoo.com/v8/finance/chart/ORCL?metrics=high?&interval=3mo&range=5y
https://query1.finance.yahoo.com/v8/finance/chart/MS?metrics=high?&interval=3mo&range=5y
https://query1.finance.yahoo.com/v8/finance/chart/TRI?metrics=high?&interval=3mo&range=5y
https://query1.finance.yahoo.com/v8/finance/chart/CRM?metrics=high?&interval=3mo&range=5y
https://query1.finance.yahoo.com/v8/finance/chart/AAPL?metrics=high?&interval=3mo&range=5y
https://query1.finance.yahoo.com/v8/finance/chart/ARMK?metrics=high?&interval=3mo&range=5y
https://query1.finance.yahoo.com/v8/finance/chart/MSFT?metrics=high?&interval=3mo&range=5y
https://query1.finance.yahoo.com/v8/finance/chart/AXP?metrics=high?&interval=3mo&range=5y
https://query1.finance.yahoo.com/v8/finance/chart/HSBC?metrics=high?&interval=3mo&range=5y
https://query1.finance.yahoo.com/v8/finance/chart/YEXT?metrics=high?&interval=3mo&range=5y
https://query1.finance.yahoo.com/v8/finance/chart/LYG?metrics=high?&interval=3mo&range=5y
https://query1.finance.yahoo.com/v8/finance/chart/WTW?metrics=high?&interval=3mo&range=5y
https://query1.finance.yahoo.com/v8/finance/chart/DVA?metrics=high?&interval=3mo&range=5y
https://query1.finance.yahoo.com/v8/finance/chart/PRSO?metrics=high?&interval=3mo&range=5y
https://query1.finance.yahoo.com/v8/finance/chart/KFY?metrics=high?&interval=3mo&range=5y
https://query1.finance.yahoo.com/v8/finance/chart/SPWR?metrics=high?&interval=3mo&range=5y
https://query1.finance.yahoo.com/v8/finance/chart/WIT?metrics=high?&interval=3mo&range=5y
https://query1.finance.yahoo.com/v8/finance/chart/CSCO?metrics=high?&interval=3mo&range=5y
```

REMOVING NAN VALUES

```
import pandas as pd

kaggle = pd.read_csv("kaggle_converted_dates.csv")


## Clean the data to drop NaN values in order to build the best regression possible
kaggle.dropna(subset=['location', 'work_life_balance', 'culture_values',
                      'career_opp', 'comp_benefits', 'senior_mgmt'], inplace=True)

kaggle.reset_index(drop=True, inplace=True)

kaggle.head()
```

	firm	date_review	job_title	current	location	overall_rating	work_life_balance	culture_values	diversity_inclusion	career_opp	...
0	McDonald-s	2020-09-16	Crew Member	Former Employee, less than 1 year	East Lansing, MI	2	3.0	3.0	5.0	1.0	...
1	McDonald-s	2020-09-19	Crew Member	Former Employee, more than 1 year	Kennewick, WA	4	5.0	3.0	5.0	1.0	...
2	McDonald-s	2020-09-22	Crew Member	Former Employee, less than 1 year	Auckland, Auckland	3	5.0	3.0	5.0	2.0	...
	McDonald-		Crew	Former Employee	Bristol,						

FORMATTING DATES

```python
import pandas as pd

file_path = 'KaggleCleaned.csv'

# Read the entire dataset
kaggle = pd.read_csv(file_path)

# Set the reference date
reference_date = pd.to_datetime('1900-01-01')

# Convert the 'date_review' column to datetime using the reference date
kaggle['date_review'] = pd.to_datetime(kaggle['date_review'], unit='D', origin=reference_date)

# Export the DataFrame to a new CSV file
output_file_path = 'kaggle_converted_dates.csv'
kaggle.to_csv(output_file_path, index=False)

# Display the DataFrame
print(kaggle.head())
```

	firm	date_review	job_title
0	McDonald-s	44088	Crew Member

	firm	date_review	job_title
0	McDonald-s	2020-09-16	Crew Member

EXPLORING THE DATA

DECISION TREE ANALYSIS

Purpose: Determine Accuracy of Models

DECISION TREE 1 ACCURACY SCORE

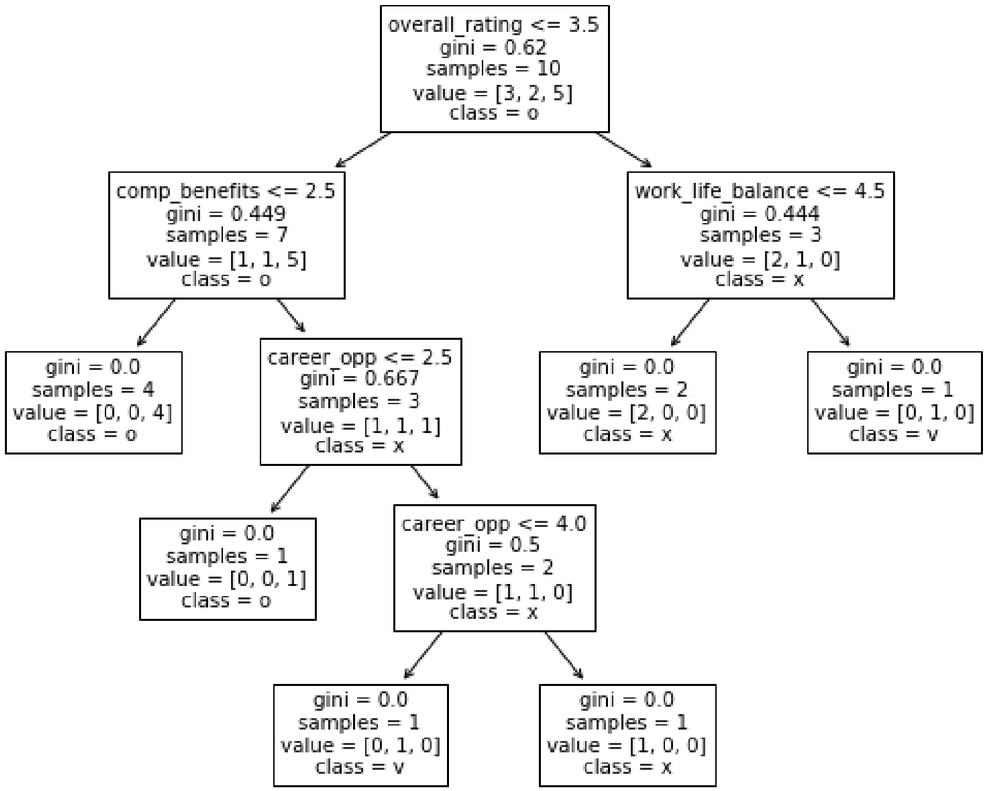
78.4%

DECISION TREE 2 ACCURACY SCORE

76.4%

PRESCRIPTIVE

RECOMMENDATIONS DERIVED



V = POSITIVE | R = MILD | X = NEGATIVE

TAKEAWAYS & RELEVANCE

- High overall rating leads to positive
- High W/L balance is more likely to have a positive review
- Low W/L balance leads to negative review
- Low career opportunities leads to negative review

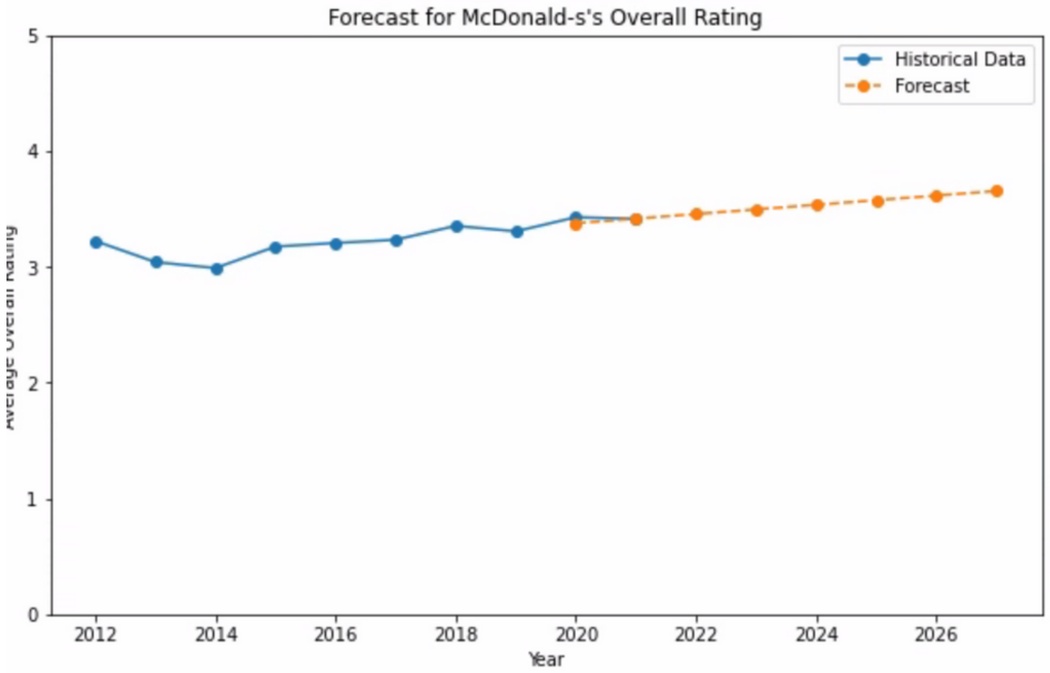
EMPLOYEES VALUE BALANCE AND OPPORTUNITY

PREDICTIVE

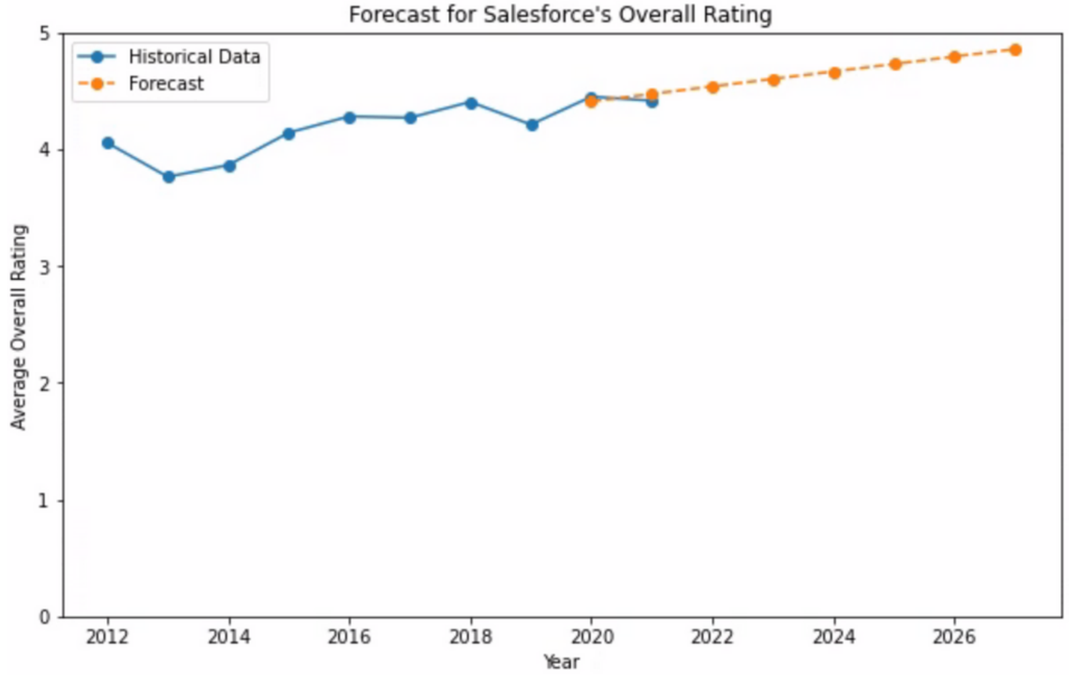
INDENTIFYING TRENDS

TIME SERIES ANALYSIS

MCDONALDS



SALESFORCE



Time Series

PREDICTIVE

RELEVANCE

TIME SERIES ANALYSIS

MCDONALDS

16% CAGR

THE CX LEAD, 2023

GREATER THAN

SALESFORCE

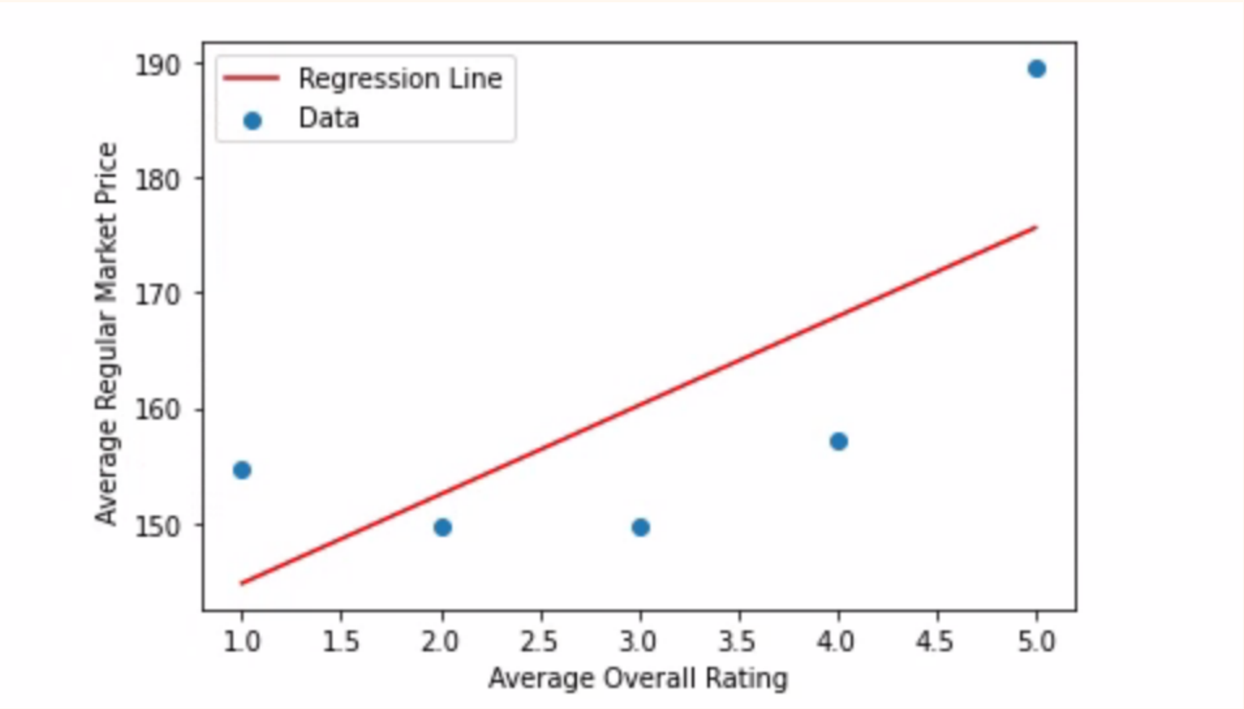
24.4% CAGR

THE CX LEAD, 2023

DETERMINING CORRELATION

REGRESSION ANALYSIS

R-SQUARED = .535

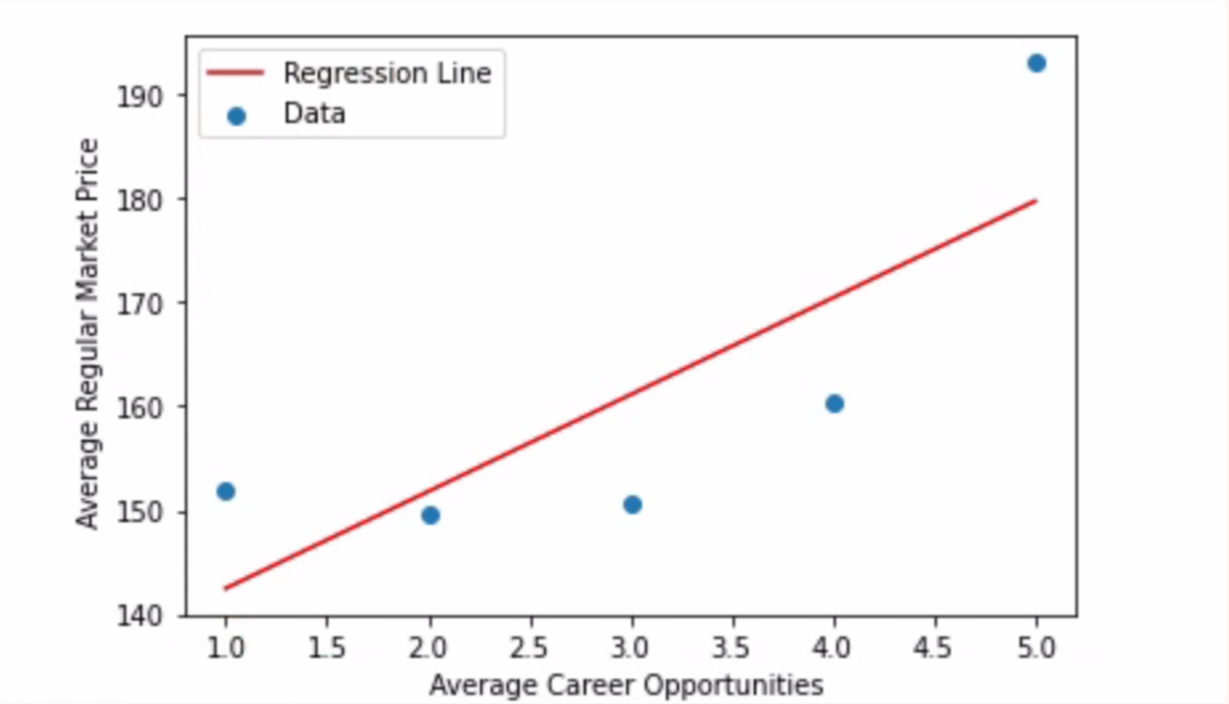


IV: OVERALL RATING| DV: REGULAR MARKET PRICE

DETERMINING CORRELATION

REGRESSION ANALYSIS

R-SQUARED = .643

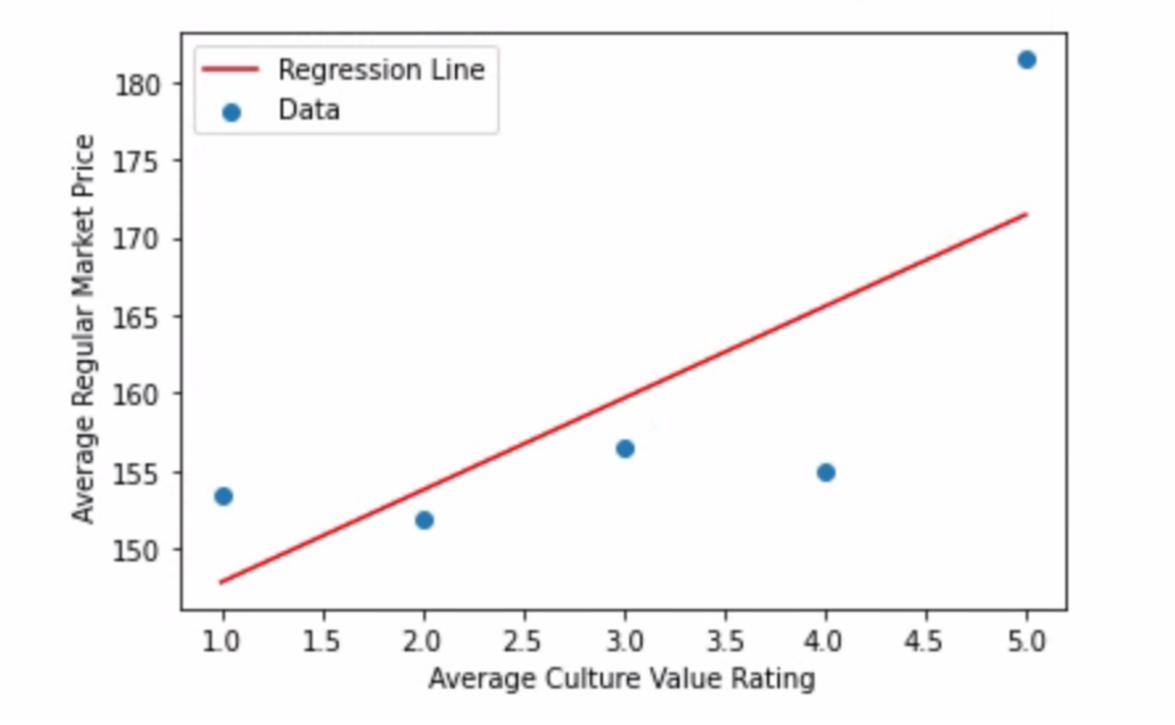


IV: CAREER OPPORTUNITIES | DV: REGULAR MARKET PRICE

DETERMINING CORRELATION

REGRESSION ANALYSIS

R-SQUARED = .575

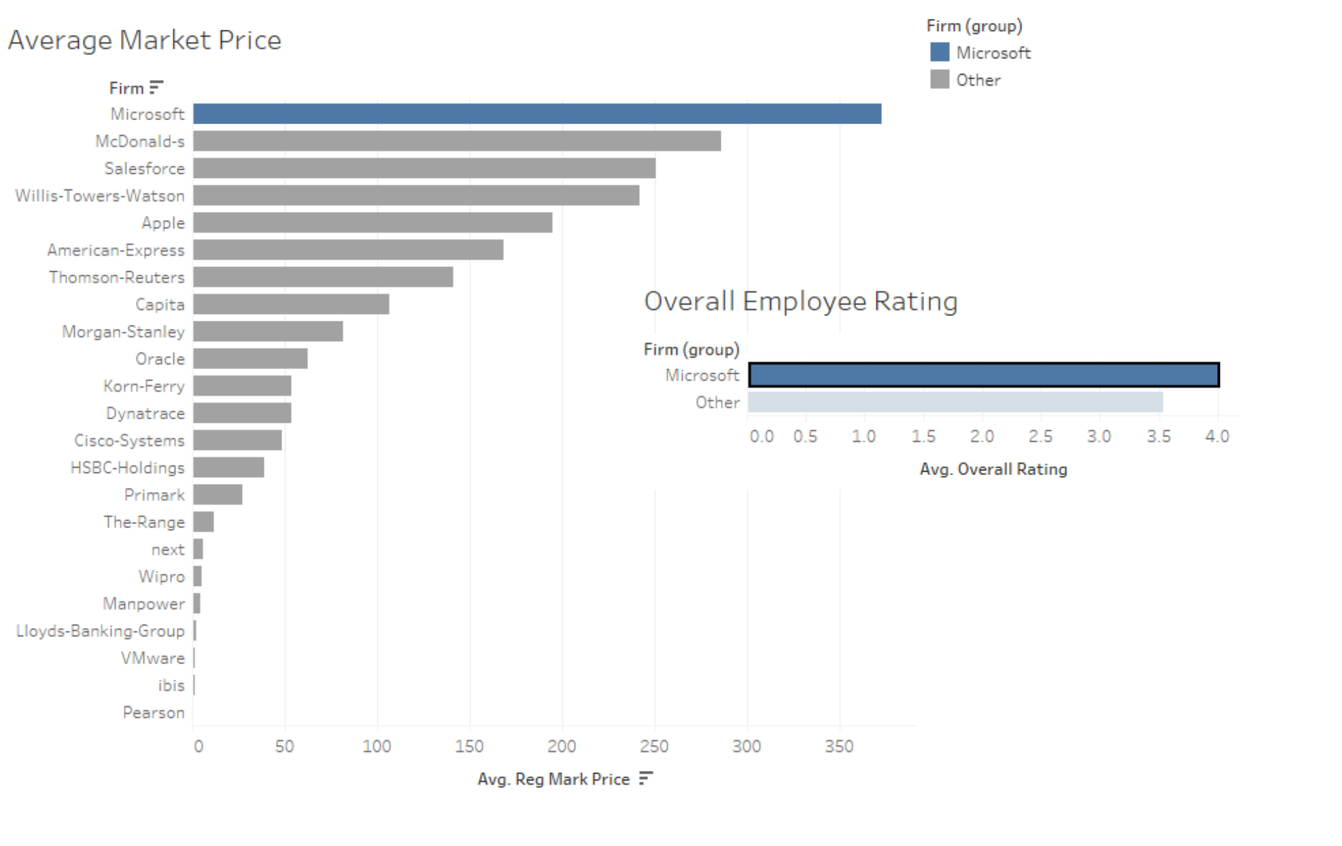


IV: CULTURE VALUES| DV: REGULAR MARKET PRICE

INDENTIFYING TRENDS

OVERALL RATING ANALYSIS

MICROSOFT - HIGHEST MARKET PRICE



VISUALIZATION FINDING:

MICROSOFT HAS A HIGHER AVERAGE EMPLOYEE RATING THAN ALL OTHER COMPANIES

CHALLENGES

CLEANING DATASETS

- Appending tickers Into Kaggle Dataset
- Pulling Market Price from Yahoo Finance

REGRESSION MODEL

- Determing relevance of the regression model
- Choose IV and DV variables

TIME SERIES

- Looping the program through every firm In the dataset
- Projecting avg rating of firms year over year

THE TAKEAWAY

THE PROBLEM

Should companies invest more into caring about their employees and things that are important to them because it positively impacts their financial performance and stock price?

OUR CONCLUSION

YES.

Through our analysis we found that companies that have higher overall employee satisfaction tend to do better financially and retain more employees.

POTENTIAL FURTHER STEPS

- Private companies
- Industry specific analysis
- Internal studies
- Google review crawl

PURPOSE: LOOK INTERNALLY AT COMPANIES AND WHAT THEY DO WELL OR DO POORLY

THANK *You*