Final Project Report

IST 718 – Professor Lando

Amanda Norwood

TABLE OF CONTENTS

SUMMARY OF FINDINGS & RECOMMENDATIONS

Gender, degree, and profession have a significant impact on affordability in Los Angeles. Men are more likely to out earn women across all degree types. Men are also more likely to out earn women in most industries as well. There is significant correlation in pay and education level as well.

Overall, 153 out of 155 professions cannot afford a home on a single salary. Housing affordability by zip code is associated to crime as well. In higher priced zip codes, crime and weapon rates taper off. In more affordable zip codes, the crime and weapon rates are close to double.

It is recommended for potential migrants to research pay based on their profession before they decide to move to Los Angeles. It is also important to factor in living with others if safety is a priority.

SPECIFICATION

PROBLEM

The goal of the analysis is to provide a method for potential L.A. migrants with information regarding the housing affordability and salary based on their gender, profession, or education for them to make informed decisions on their move. We also want to predict zip codes where people can live based on their demographic data.

HYPOTHESIS

Hypothesis: Gender, degree, or profession do not have an impact on housing affordability.

Alternative hypothesis: Gender, degree, or profession DO have an impact on housing affordability.

DATA

EDUCATION DATA – 126 ROWS, 15 COLUMNS

	ID Gender	Gender	ID Group	Group	ID Year	Year	ID State	State	Total Population	Total Population MOE Appx	Average Wage	Average Wage Appx MOE	Group ID	Age Range	Percentage
0	2	Female	Associates Degree	Associates Degree	2020	2020	04000US06	California	1161339	4012.757389	22279.847898	2835.671997	7	1	55.261437
1	2	Female	Bachelors Degree	Bachelors Degree	2020	2020	04000US06	California	2994902	7539.885565	40526.050930	2986.719653	8	1	52.004361
2	2	Female	Graduate Degree	Graduate Degree	2020	2020	04000US06	California	1747439	3818.350081	46913.297071	6708.984104	9	1	50.422991
3	2	Female	High School or Equivalent	High School or Equivalent	2020	2020	04000US06	California	2695919	6122.901609	17423.896492	2326.322214	5	1	49.180740
4	2	Female	No Schooling	No Schooling	2020	2020	04000US06	California	393477	1192.915610	8094.061415	7463.503355	1	1	53.622708

Relevant columns include gender, degree type, reported year, and average wage.

MEDIAN EARNINGS – 208 ROWS, 11 COLUMNS

	ID Gender	Gender	ID Industry Group	Industry Group	ID Year	Year	Median Earnings by Industry and Gender	Median Earnings by Industry and Gender Moe	Geography	ID Geography	Slug Geography
0	0	Male	1	Agriculture, Forestry, Fishing & Hunting, & Mi...	2020	2020	27462	5706	Los Angeles, CA	16000US0644000	los-angeles-ca
1	0	Male	2	Construction	2020	2020	31123	358	Los Angeles, CA	16000US0644000	los-angeles-ca
2	0	Male	3	Manufacturing	2020	2020	39168	1360	Los Angeles, CA	16000US0644000	los-angeles-ca
3	0	Male	4	Wholesale Trade	2020	2020	40638	1416	Los Angeles, CA	16000US0644000	los-angeles-ca
4	0	Male	5	Retail Trade	2020	2020	28440	991	Los Angeles, CA	16000US0644000	los-angeles-ca

Relevant columns include gender, occupation type, year, and median earnings.

OCCUPATIONS – 198 ROWS, 23 COLUMNS

	ID Group	Group	ID Subgroup	Subgroup	ID Occupation	Occupation	ID Year	Year	ID State	State	...	Geography	ID Geography	Slug Geography	Median Earnings
0	0	Management, Business, Science, & Arts Occupations	0	Management, Business, & Financial Occupations	0	Management Occupations	2020	2020	04000US06	California	...	Los Angeles, CA	16000US0644000	los-angeles-ca	147944
1	0	Management, Business, Science, & Arts Occupations	0	Management, Business, & Financial Occupations	1	Business & Financial Operations Occupations	2020	2020	04000US06	California	...	Los Angeles, CA	16000US0644000	los-angeles-ca	136895
2	0	Management, Business, Science, & Arts Occupations	1	Computer, Engineering, & Science Occupations	2	Computer & Mathematical Occupations	2020	2020	04000US06	California	...	Los Angeles, CA	16000US0644000	los-angeles-ca	151426
3	0	Management, Business, Science, & Arts Occupations	1	Computer, Engineering, & Science Occupations	3	Architecture & Engineering Occupations	2020	2020	04000US06	California	...	Los Angeles, CA	16000US0644000	los-angeles-ca	161213
4	0	Management, Business, Science, & Arts Occupations	1	Computer, Engineering, & Science Occupations	4	Life, Physical, & Social Science Occupations	2020	2020	04000US06	California	...	Los Angeles, CA	16000US0644000	los-angeles-ca	123577

Relevant columns include occupation type, year, and median earnings.

HOME PRICE BY ZIP CODE – 275 ROWS, 8 COLUMNS

	Zip Code	City / Community	2023*	2022	2021	2020	2019	2018
0	90001	Los Angeles (South Los Angeles), Florence-Graham	540635.0	555034.0	508199	465268	425927	404863
1	90002	Los Angeles (Southeast Los Angeles, Watts)	539102.0	552950.0	518364	463726	421153	398477
2	90003	Los Angeles (South Los Angeles, Southeast Los ...	558141.0	577163.0	542654	480137	435456	413903
3	90004	Los Angeles (Hancock Park, Rampart Village, Vi...	1767062.0	1824797.0	1722525	1578331	1453209	1477493
4	90005	Los Angeles (Hancock Park, Koreatown, Wilshire...	1791046.0	1880689.0	1774383	1602551	1487273	1504952

Relevant columns include the zip code, city/community, and the home prices across the years.

CRIME – 710549 ROWS, 31 COLUMNS

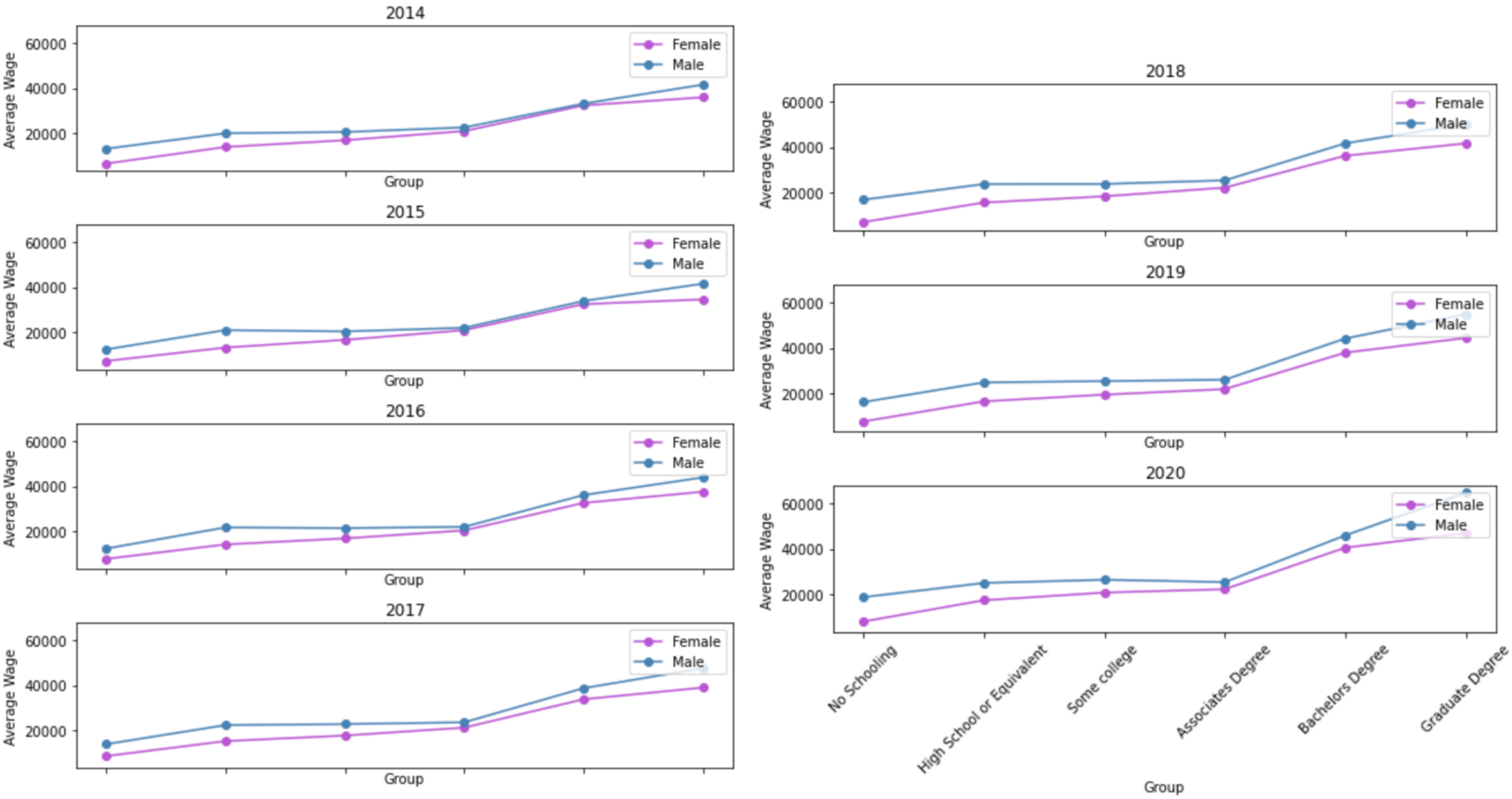
	Zip Code	City / Community	2023*	2022	2021	2020	2019	2018
0	90001	Los Angeles (South Los Angeles), Florence-Graham	540635.0	555034.0	508199	465268	425927	404863
1	90002	Los Angeles (Southeast Los Angeles, Watts)	539102.0	552950.0	518364	463726	421153	398477
2	90003	Los Angeles (South Los Angeles, Southeast Los ...	558141.0	577163.0	542654	480137	435456	413903
3	90004	Los Angeles (Hancock Park, Rampart Village, Vi...	1767062.0	1824797.0	1722525	1578331	1453209	1477493
4	90005	Los Angeles (Hancock Park, Koreatown, Wilshire...	1791046.0	1880689.0	1774383	1602551	1487273	1504952

Relevant columns include the zip code, crime, and weather a weapon was used.

OBSERVATIONS

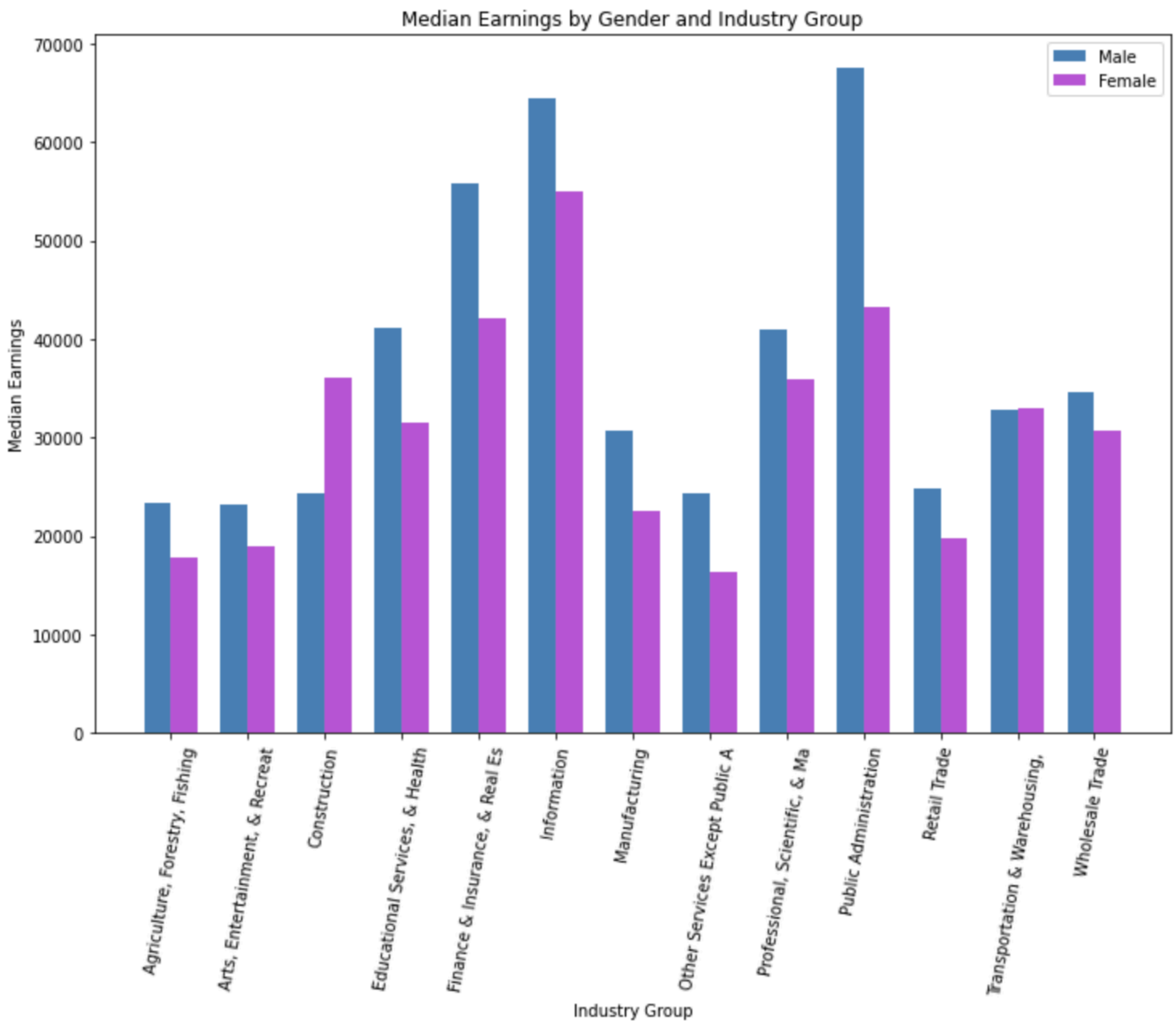
PAY BY GENDER & DEGREE TYPE

The data provides average salary information by degree and gender from 2014 – 2020. On average, women have earned less than men across all years and degree types.


2014
2015
2016
2017
2018
2019
2020
Average Wage
60000
40000
20000
Female
Male
Group
No Schooling
High School or Equivalent
Some college
Associates Degree
Bachelors Degree
Graduate Degree

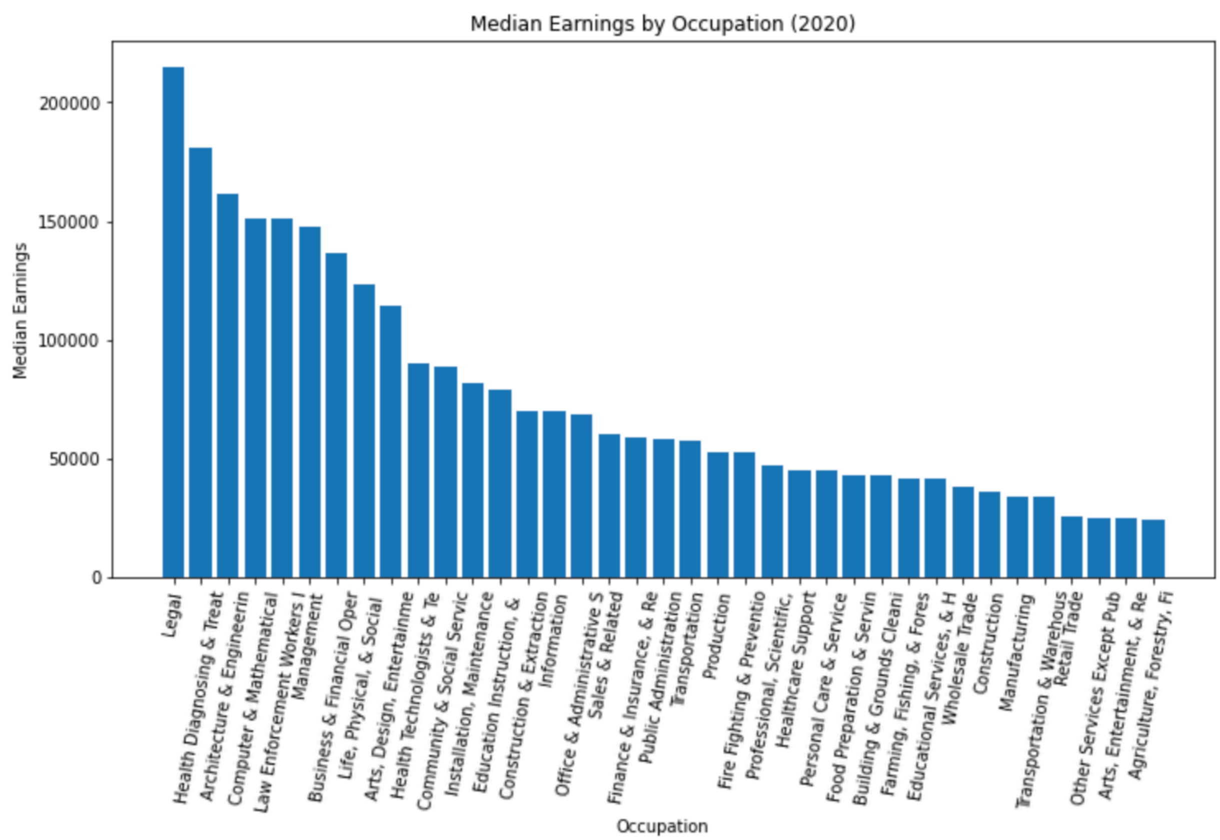
MEDIAN EARNINGS BY GENDER AND INDUSTRY GROUP

Men generally out earn women except for in the construction and transportation & warehouse field.



OCCUPATIONS

The occupations dataset doesn't include gender—only occupation and salary. Legal and health care related fields earn the most. Arts & entertainment and agriculture earn the least.



ANALYSIS

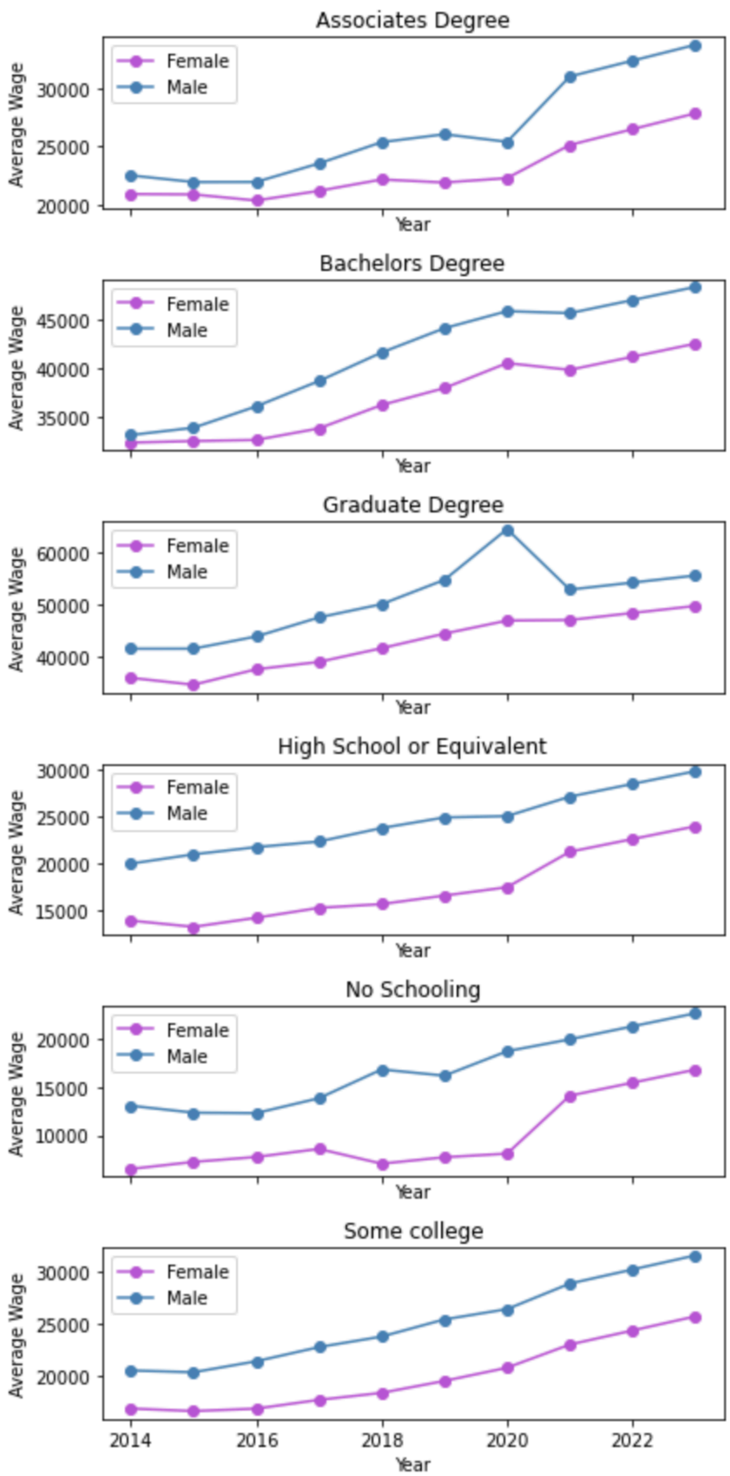
FORECASTING SALARY WITH LINEAR REGRESSION

In our original dataset, the salary only went to 2020. We used linear regression to predict the values for 2021 through 2023.

forecast

```python
data = educationpred.copy()
features = ['Gender', 'Group', 'Year']
target = 'Average Wage'
numerical_features = []
categorical_features = []
for feature in features:
    if data[feature].dtype == 'object':
        categorical_features.append(feature)
    else:
        numerical_features.append(feature)
preprocessing = ColumnTransformer(
    transformers=[
        ('num', StandardScaler(), numerical_features),
        ('cat', OneHotEncoder(handle_unknown='ignore'), categorical_features)
    ])

# split the data into training and test
X = data[features]
y = data[target]
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

model = LinearRegression()

pipeline = Pipeline(steps=[
    ('preprocessing', preprocessing),
    ('model', model)
])
# train the model
pipeline.fit(X_train, y_train)
# predict average wage
years = range(2021, 2024)
categories = data['Group'].unique()
genders = data['Gender'].unique()
predicted_data = []
for year in years:
    for category in categories:
        for gender in genders:
            new_data = pd.DataFrame({
                'Gender': [gender],
                'Group': [category],
                'Year': [year]
            })
            new_data_processed = pipeline.named_steps['preprocessing'].transform(new_data)

            predicted_wage = pipeline.named_steps['model'].predict(new_data_processed)[0]
            predicted_data.append({
                'Gender': gender,
                'Group': category,
                'Year': year,
                'Average Wage': predicted_wage
            })

predicted_df = pd.DataFrame(predicted_data)
```

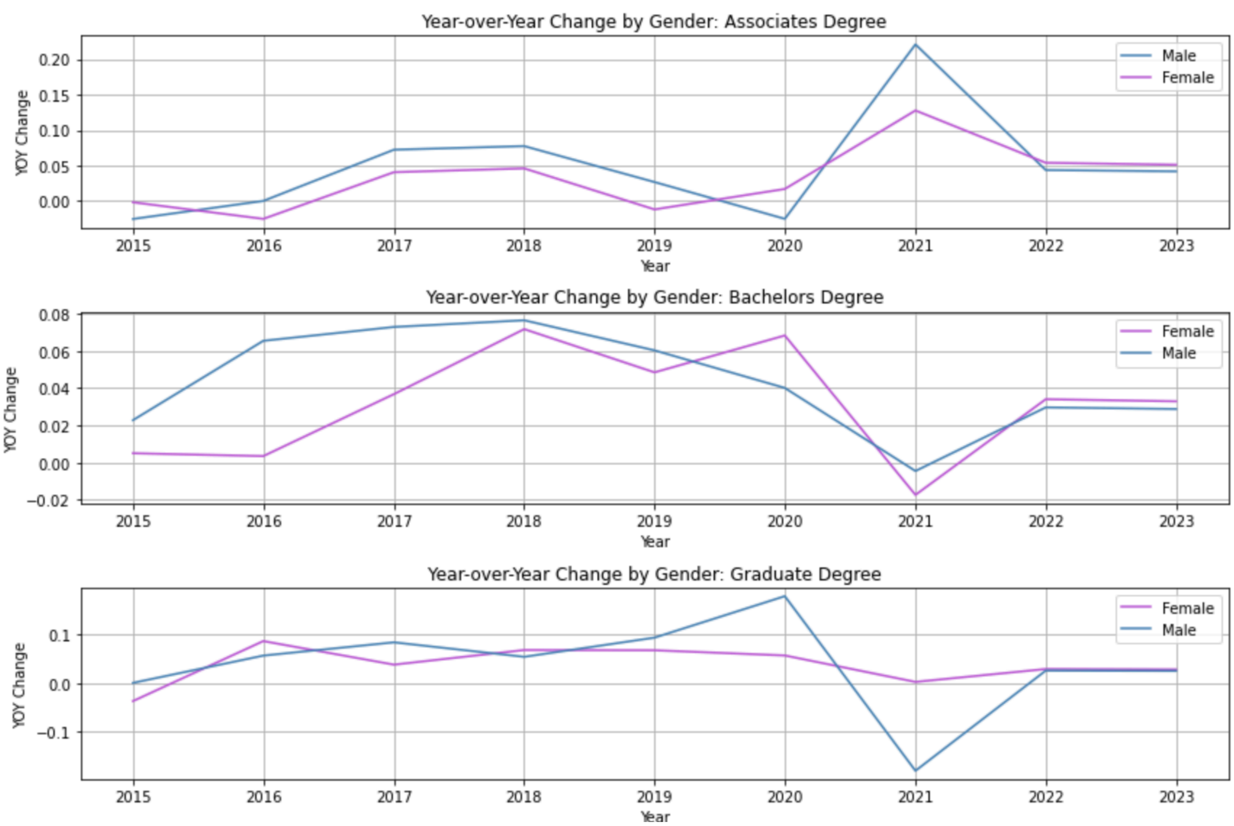
Here are the results after the prediction:



CALCULATING YEAR OVER YEAR CHANGE

Calculate year over year change to visually graph market change.

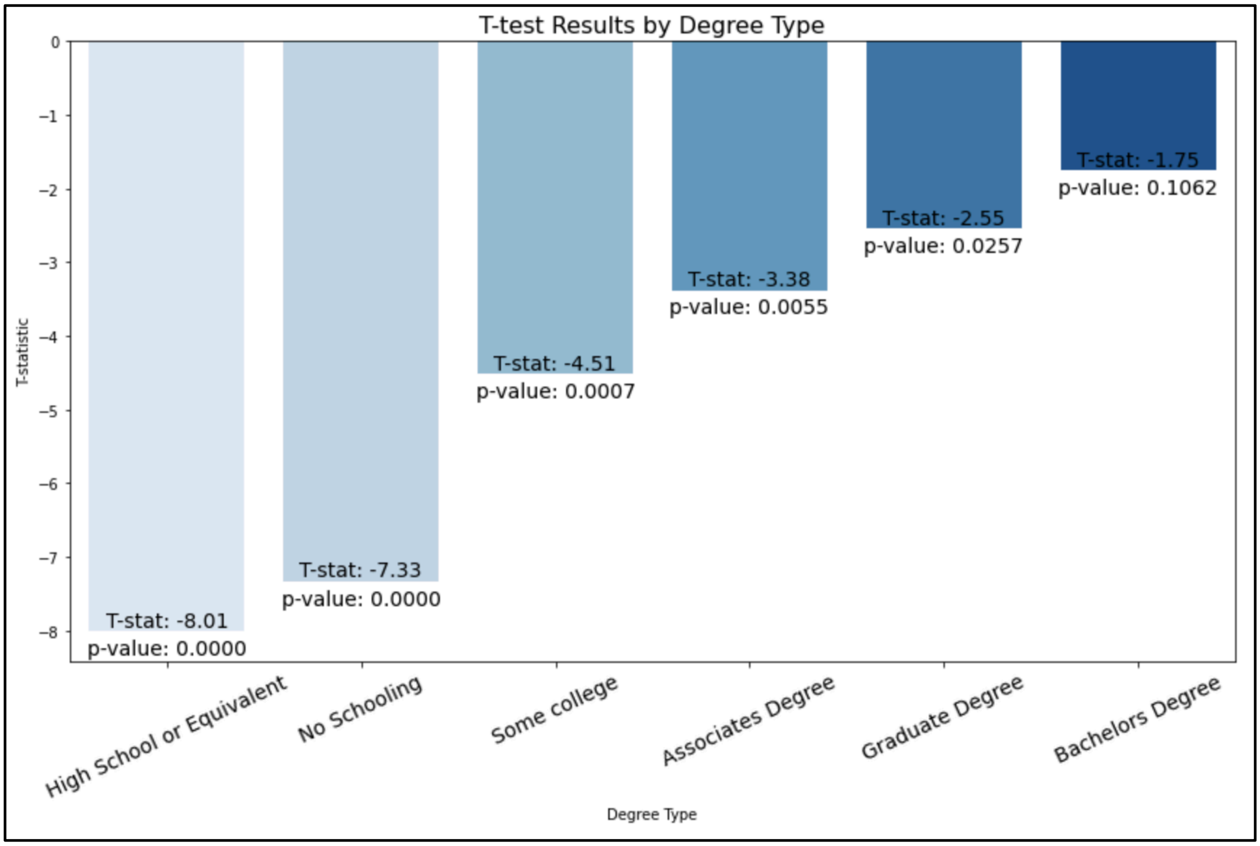
```
# prep to graph YOY change
educationYOY = salaryprediction[['Gender','Group','Year','Average Wage']]

df = educationYOY
df['Year'] = pd.to_datetime(df['Year'], format='%Y')
df.sort_values(['Year', 'Group'], inplace=True)

# calculate YOY change
df['YOY Change'] = df.groupby(['Gender', 'Group'])['Average Wage'].pct_change()
# drop na values since the first year will be null
df.dropna(subset=['YOY Change'], inplace=True)
df.head(1)
```



T-TEST BY GENDER AND DEGREE TYPES

The T-test measures the difference between the means of two groups. In this case, we are comparing the pay between genders by degree types. When looking at the t-test results in absolute values, it indicates a larger difference associated between the means. There is a p-value associated with the t-test as well. It indicates the likelihood of observing the difference in means by chance alone.

According to t-test results, HS degree and no schooling show the largest variability in pay between men and women. Each degree type has a p-value less than 0.05 except for bachelor’s degree. This means that bachelor’s degree does not show difference in pay for men and women, while the rest do.

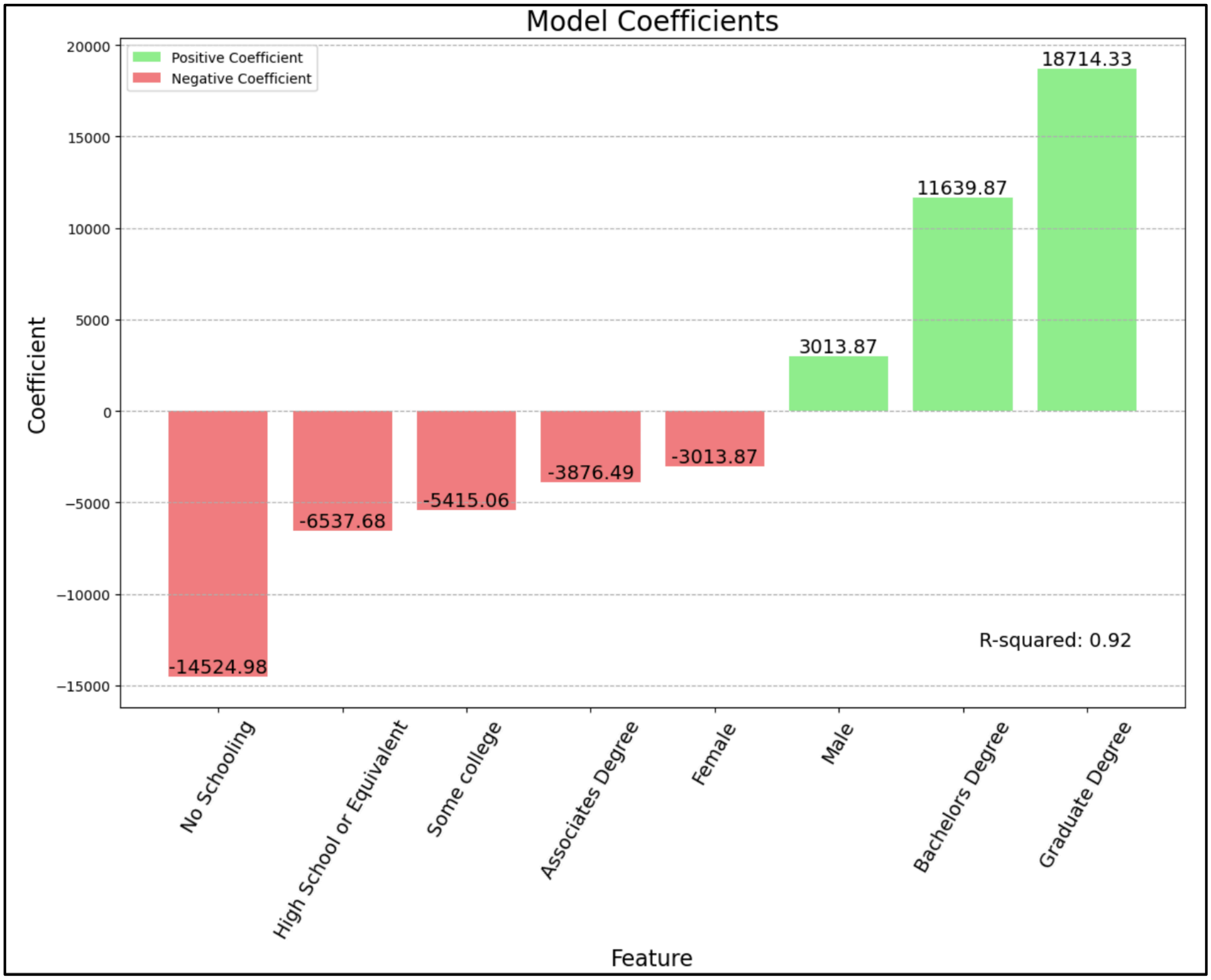


LINEAR REGRESSION

For linear regression, our independent variables are gender and degree type. And our dependent variable is the average wage.

R squared is 92%, which indicates is a good fit in explaining the variability in wage based on gender and degree type. The coefficients represent the expected change in wage depending on the independent variable.

Females, on average, are earn an average of roughly $3000 less than males. And those with graduate degrees, are more likely to earn, on average, $18.7k more.



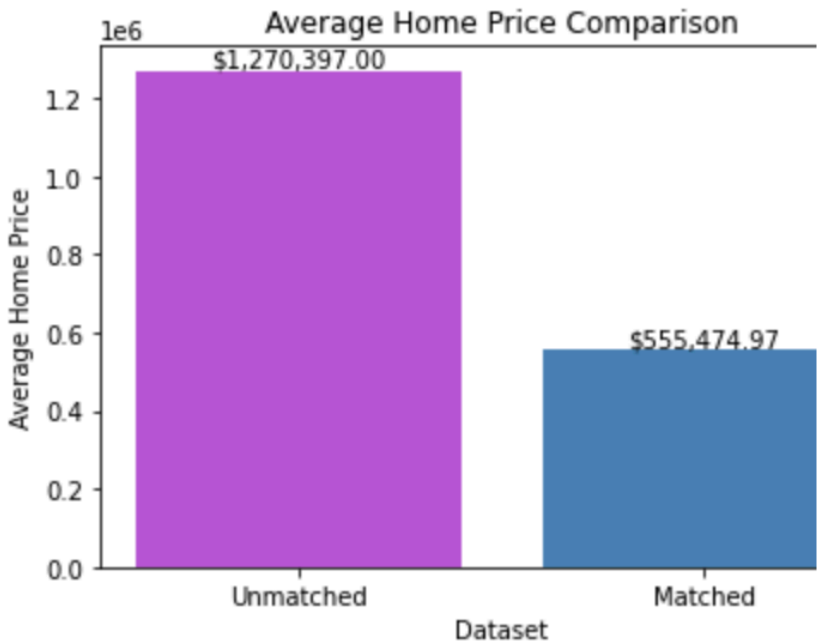
HOUSING AFFORDABILITY AND CRIME ZIP CODE MATCH

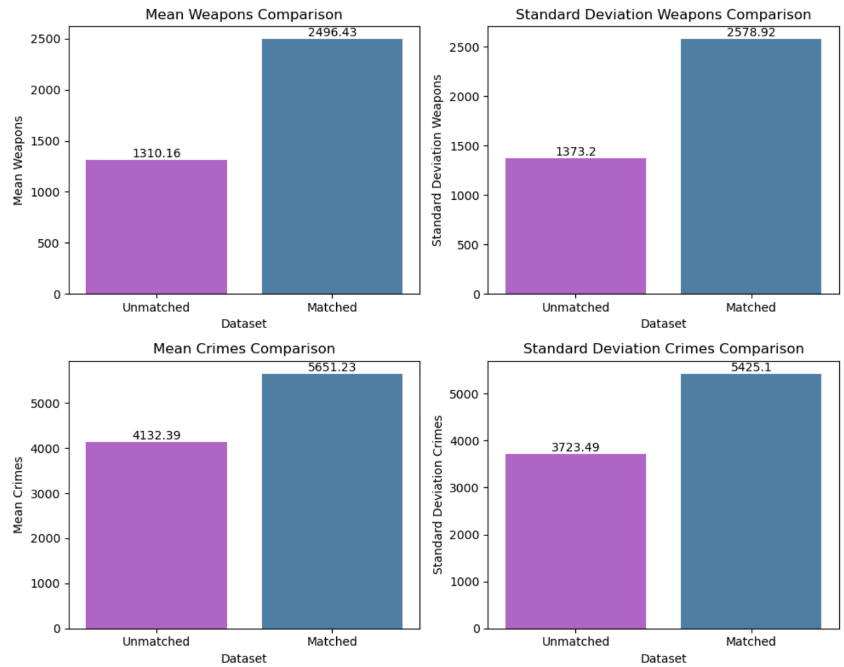
We determined which professions could afford a home in L.A. County by calculating affordability based on three times their salary.

We compared the rates of crime between the zip codes where the matched professions could afford a home and the zip codes where they couldn't. This allowed us to understand if there was any difference in the levels of crime between these two groups.

We can see that unmatched homes are over double the cost of matched homes. And their crime and weapon rates are much lower as well.

Average home price for unmatched homes is double the matched homes. Likewise, crime and weapon rates are also doubled.





After analysis, only two professions could afford a home on a single salary based on their 3x income:

Professions that can afford a home on a single salary:
Health Diagnosing & Treating Practitioners & O...
Legal

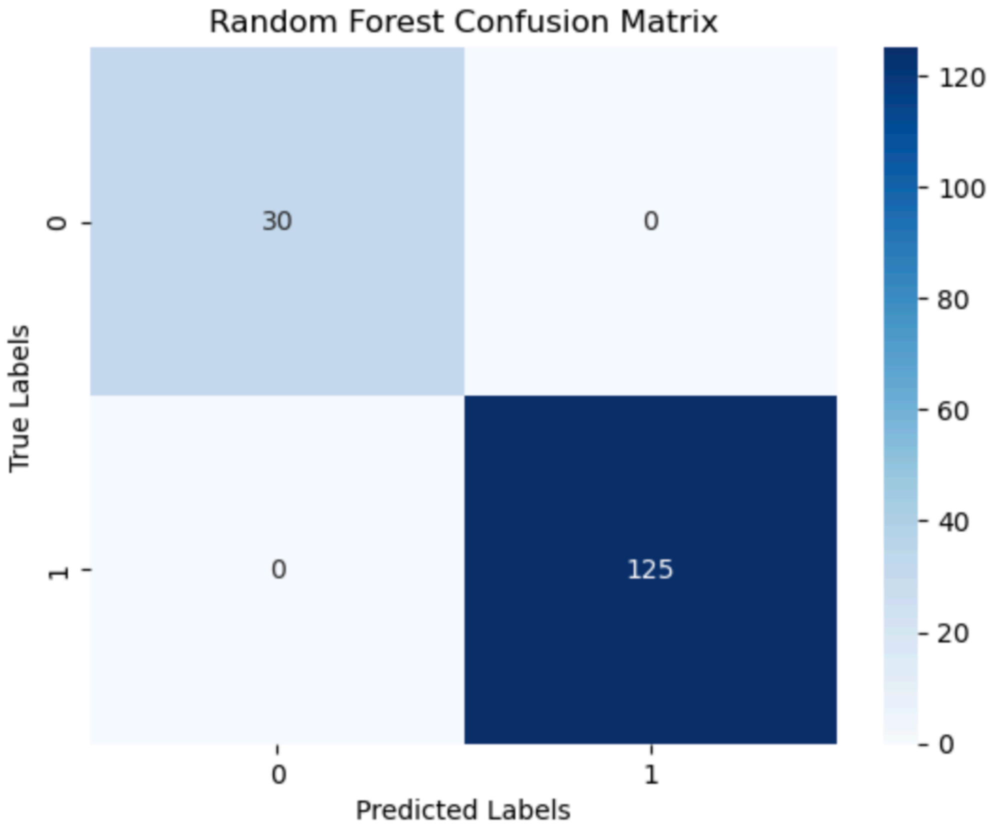
And these professions would need a roommate or someone to purchase with them:

Professions that cannot afford a home on a single salary:
Construction & Extraction
Education Instruction, & Library
Management
Life, Physical, & Social Science
Food Preparation & Serving Related
Farming, Fishing, & Forestry
Fire Fighting & Prevention, & Other Protective...
Business & Financial Operations
Arts, Design, Entertainment, Sports, & Media
Computer & Mathematical
Transportation
Architecture & Engineering
Health Technologists & Technicians
Production
Office & Administrative Support
Healthcare Support
Community & Social Service
Law Enforcement Workers Including Supervisors
Sales & Related
Installation, Maintenance, & Repair
Building & Grounds Cleaning & Maintenance
Personal Care & Service

RANDOM FOREST

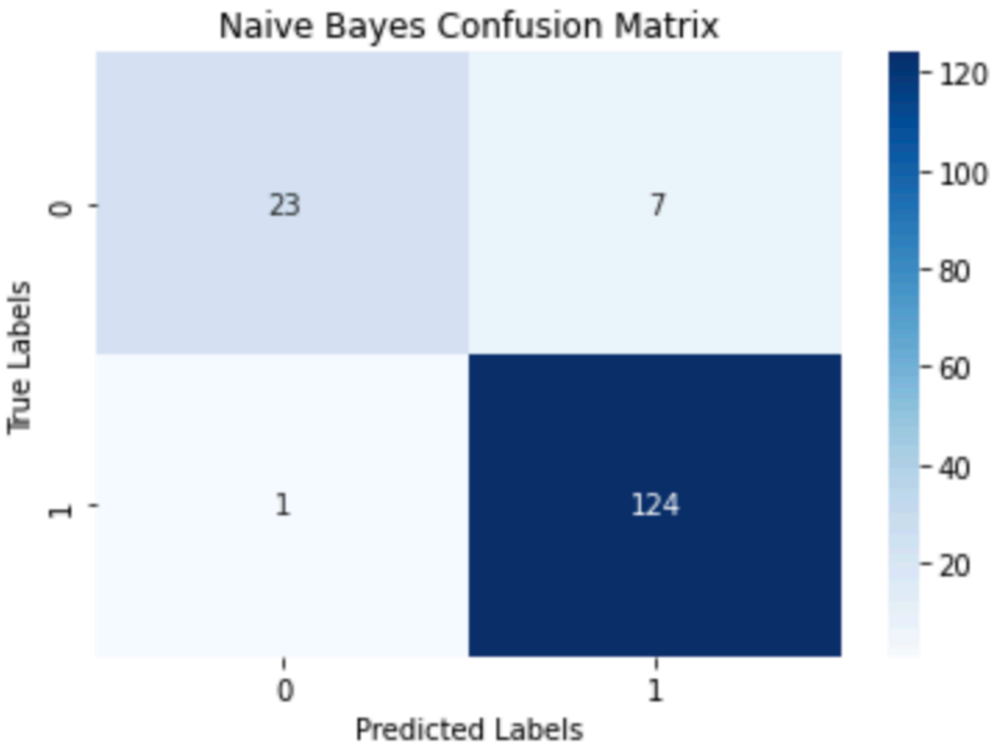
Random Forest (decision trees) was used to predict whether professions are matched or unmatched based on housing affordability.

The random forest classifier performed best with an accuracy rate of 100%. If there are other professions that could be entered into the algorithm, then it would be a great way to predict whether a profession or salary could afford a home and within which zip code.



NAÏVE BAYES

Naive Bayes (probability algorithm) were used to predict whether professions are matched or unmatched based on housing affordability and its accuracy rate is 94.84%.



RECOMMENDATIONS

Out of 155 professions, only 2 professions could afford to purchase a home in L.A. on a single salary. In addition, the t-test resulted in highlighting pay gaps between male and females across all types of education except for bachelor's degree. On average, women earn roughly $3,000 less than males. And those with graduate degrees are more likely to earn about $18,700 more.

In zip codes where professions can afford a home on a single salary, crime and weapon rates are exponentially higher and are almost doubled. The average, affordable home in 2023 costs around $550,000, while unaffordable homes average to $1,270,000.

Based on the analysis, it is concluded that gender, degree, and profession have an impact on housing affordability. It is recommended that people are aware of potential gender disparities in salary. Women are more likely to earn less across all degree types, but men also earn less than women in certain professions. Everyone should consider negotiation their salary for a fair compensation, regardless of gender.

Also, when people are deciding their degree or profession, it’s important to consider the reported average wages for the specific area. The information could provide valuable insights into income potential and help guide career choices such as potentially going back to school or changing careers.

Lastly, if safety is a top priority, then it may be beneficial to explore the option of living with roommates. This can help mitigate housing affordability challenges and provide an added later of security by sharing the cost and responsibility of a property.