Implications of Demographics on Income

STAT-S 670: Exploratory Data Analysis

Sai Ajay Vutukuri, Himani Anil Deshpande, Murali Krishna Parvataneni, Tushar Kant Samantaray

Introduction

This research aims to understand how demographics affects the median income of a group of people in major labor industries during the period 2015-2020. We analyze if there is any inequality in wage based on gender, age, and race. Understanding the wage gap and addressing this issue would highlight the group of people subjected to inequality. We utilized the CPS (Current Population Survey) to extract the demographic factors such as gender, age, race, major industries, education attainment, employment status and income for the years 2015 – 2020. The CPS collects data about employment status and other population related variables.

We divided our analysis into four sections to understand how the median income is influenced by demographic variables:

0. Loading and preprocessing of data
1. Understand how median income of a group varies based on various demographic groups, such as gender, age, race, major industries, and education attainment and whether unemployment rate affects the variation in median income.
2. Fitting a model to predict the median income based on these demographic variables.
3. Verify whether introducing interactions between the demographic variables improve our model.
4. Make predictions on the model.

0. Loading and preprocessing the data

The CPS data [1] extracted from the IPUMS has 3.7 million records for years from 2015 to 2020 which were reduced to 500K records after performing the following data preprocessing steps:

a) Filtering out the data consisting of Education = "NIU or blank" or "Missing/Unknown", Labor Force = "Not in Universe", Industry code = "0" and Race = "Blank" or "NA", According to IPUMS, NIU (Not in Universe) which means the data is missing.

b) Segmenting approximately 300 minor industries into 14 Major categories based on their industry IDs obtained from census.gov [2].

c) Categorizing 26 different races into 4 groups (White, Black, Asian, and Others)

d) Grouping 17 different education levels into 6 categories ("Less than a high school diploma", "High school graduates", "Some college, no degree", "Associate degree", "Bachelor's degree", "Advanced degree").

e) When it comes to age groupings, we've seen that the under 18 age group contains missing values, hence, we excluded those under the age of 18 from our study.

f) Post filtering out and categorizing the data we omitted the null values from the data, reducing our dataset to 523,452 rows.

After pre-processing the data, we adjusted for inflation by using the CPI value of 1999 to standardize the weekly earnings for all the years from 2015-2020 [3].

IPUMS provides two sample weights ASECWT and EARNWT, which is to adjust for being under-represented or over-represented in a sample, for demographic variables and weekly earnings respectively.

1. **Understand how median income of a group varies based on various demographic groups, such as gender, age, race, major industries, and education attainment and whether unemployment rate affects the variation in median income.**

Unemployment trend over the years

The average unemployment rate is approximately 5.3% across all industries for the years between 2015 and 2020, which aligns with the real-time data provided by the Bureau of Labor Statistics [4]. Since people in Active Duty are excluded from CPS data as it focuses on the civilian population which can be seen here, Hence, we get a 0% employment rate for the Armed Forces. Therefore, in the further analysis we will not be including the Armed Forces for better results.

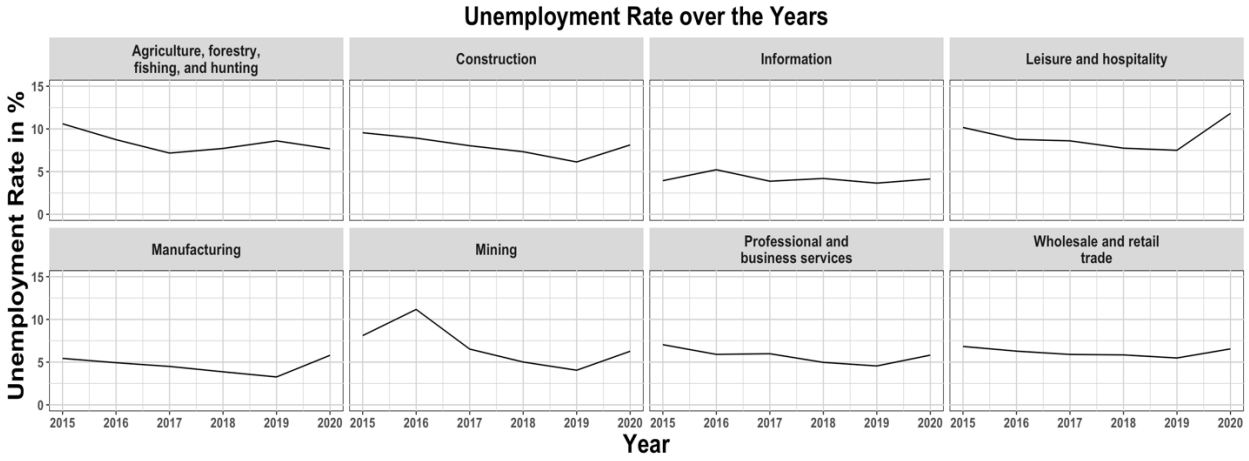


Figure 1

Figure 1 shows how the unemployment rate has varied over time in 8 major industry sectors from 2015 to 2020. In general, unemployment rates decreased from 2015 to 2019 across all sectors, except the mining and information industries, which experienced a 7% and 2% increase in unemployment rates respectively, in 2016. From 2019 to 2020, every sector experienced an increase in unemployment, possibly due to the COVID-19 Pandemic, except agriculture, forestry, fishing, and hunting, which saw a downwards trend in unemployment. The leisure and hospitality sectors appear to have been impacted the most of the Pandemic's effects, followed by manufacturing, mining, and construction. The information, financial, and public administration sectors were allegedly not impacted (the latter two not included in the graph) by the epidemic and experienced no substantial changes in unemployment rates. As there were not many significant trends and insights in Educational and health services, other services, Transportation and utilities, and public sector, they were left out of our analysis.

Change in income over the period

We began by looking at the change in median income across all industries throughout the period between 2015 and 2020 as shown in Figure 2.

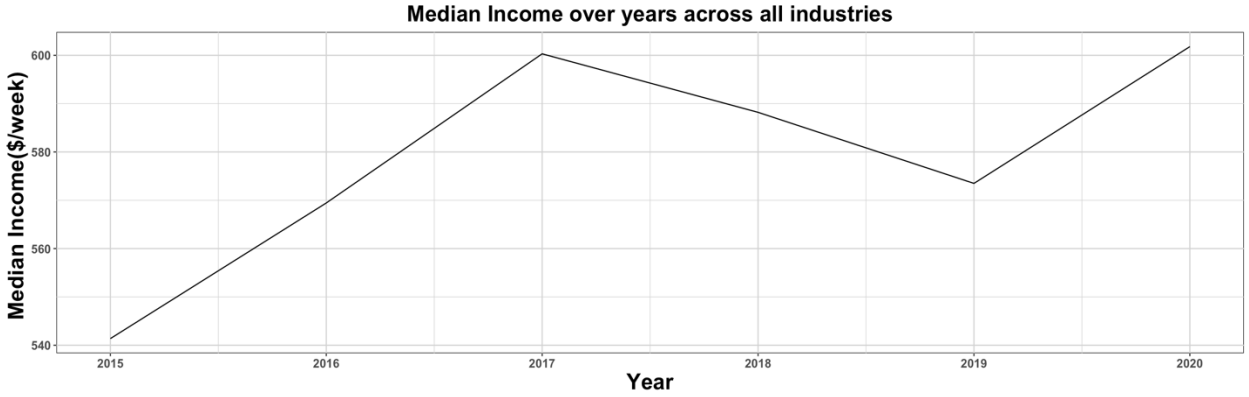


Figure 2

We used EARNWT and the CPI dollar value of 1999 to adjust the income per week rates for inflation, and we employed weighted means to accomplish this by the following formula.

$$\boldsymbol{MEDIAN\_INCOME} = \text{ weighted. median (Weekly earning * CPI99, Earn weight) [2]}$$

The median income in Figure 2 shows an increasing trend over the years. There was a notable increase in income per week from 540 $/week to 600 $/week from 2015 to 2017. We saw a dip in income from 2017 to 2019, but it has since been increasing. This could possibly be due availability of more low playing job openings during those years. Since we saw that the unemployment rate increased during the pandemic, but weekly incomes continued to rise. It could possible that those with low-paying positions were more likely to lose their jobs, which could be the reason for the unexpected increase in income during the pandemic. Let's examine if the pattern is consistent across all industries.

Overall median income across each industry

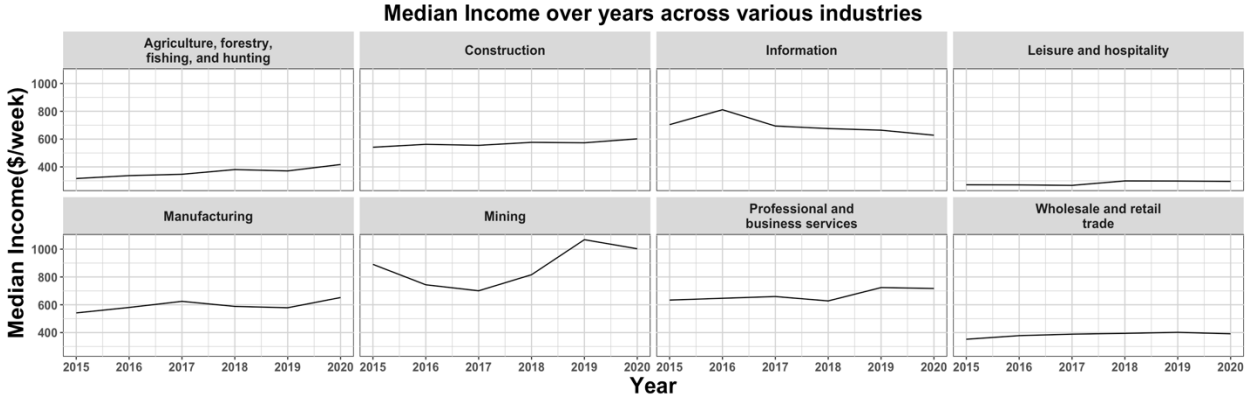


Figure 3

Figure 3 depicts how the median income varies over time and in various labor industries. The industry of leisure and hospitality showcases the least earned income among all the other industries, closely followed by wholesale and retail trade. From 2015 to 2020, earnings per week in agriculture, forestry, fishing, and hunting, construction, manufacturing, and professional and business services are seen to be rising, while earnings in the Information industry peaked in 2016, they have since declined until 2017 and have stayed stable. Overall, the mining business appears to be paying more, however there is a lot of fluctuation across time with the lowest dip in 2017 and the highest peak in 2019 before the pandemic. Even though the unemployment rate increased across the industries during the pandemic of 2019 (see Figure1), we do not detect a drop in median weekly earnings across industries, except for the mining industry. Furthermore, we investigate whether persons of various demographics follow the same pattern across industries through time or if one group of a certain demography is affected more than the other.

Change in Median Income by Gender over years

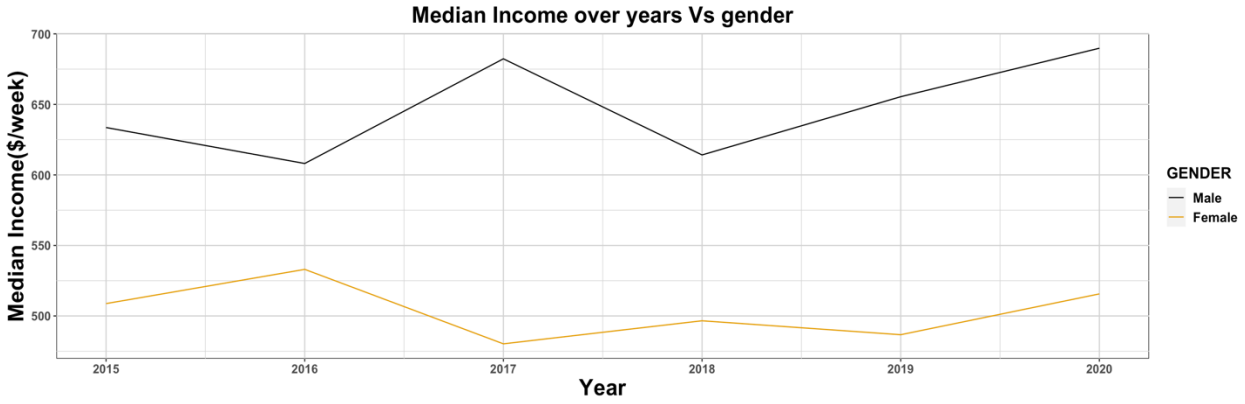


Figure 4

Despite minor variations, the overall median income of men has grown. We see a substantial wage discrepancy between males and females, with females experiencing a significant drop in income throughout the 2016-2017 period, while during the same period the wage for males have seen to reach its peak. From 2016 to 2018, there was an increase followed by a fall for males, and the weekly income has since been increasing. The rate of increase for both the genders throughout the 2019-2020 period is modest.

To fully comprehend the salary disparity between men and women, a more in-depth analysis is required, one that compares salaries depending on job location, demographics, and job description. Going further, we observe if we find the same trend across all industries for both the genders.

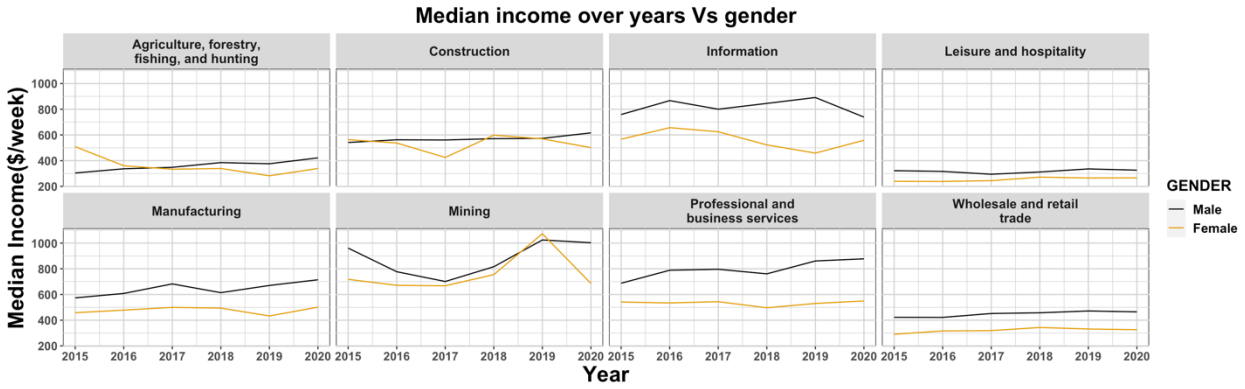


Figure 5

The wage gap between men and women is smaller in the agriculture, forestry, fishing, and hunting, construction, leisure and hospitality, and mining industries, but it is the largest in the information industry, followed by professional and business services, manufacturing, and wholesale and retail trade. As a result, the salary disparity is not a consistent trend across all industries.

Change in Median Income by Race over years

From figure 6, we notice that the Asian-Pac-Islander group seems to be the most earning group across all the years and see further increase in median income over the years, while the American Indian-Aleut-Eskimo group seems to be earning the least. The overall trend for all the groups shows an increase over the years, with the 'Others' seeing the most increase, while the American Indian-Aleut-Eskimo shows the least increment. We will look at the median income trend across industries.

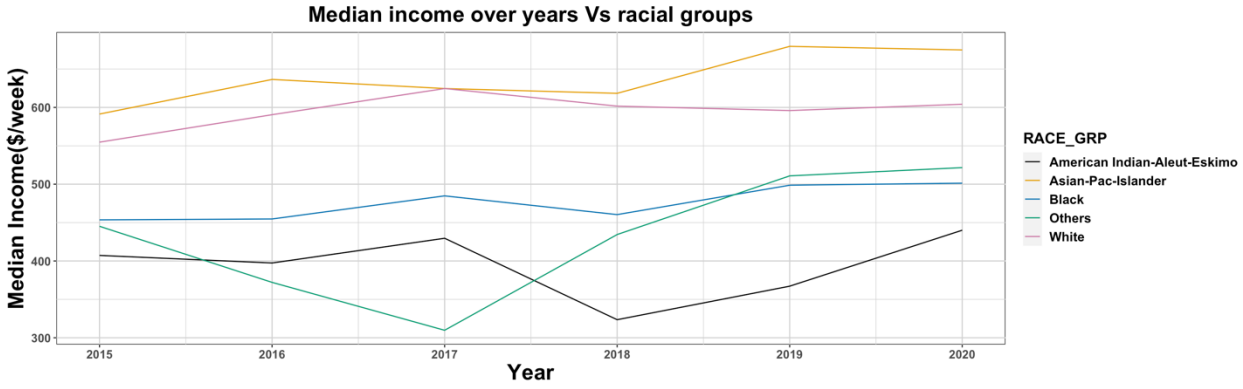


Figure 6

From figure 7, we notice that Asian-Pac-Islander group seems to be earning the most in Information, Manufacturing, Mining and Professional and business services industries, however, there is no significant difference in other industries based on racial groups. The leisure and hospitality industry, as well as the wholesale and retail industry, have the smallest disparity in median salaries between races.

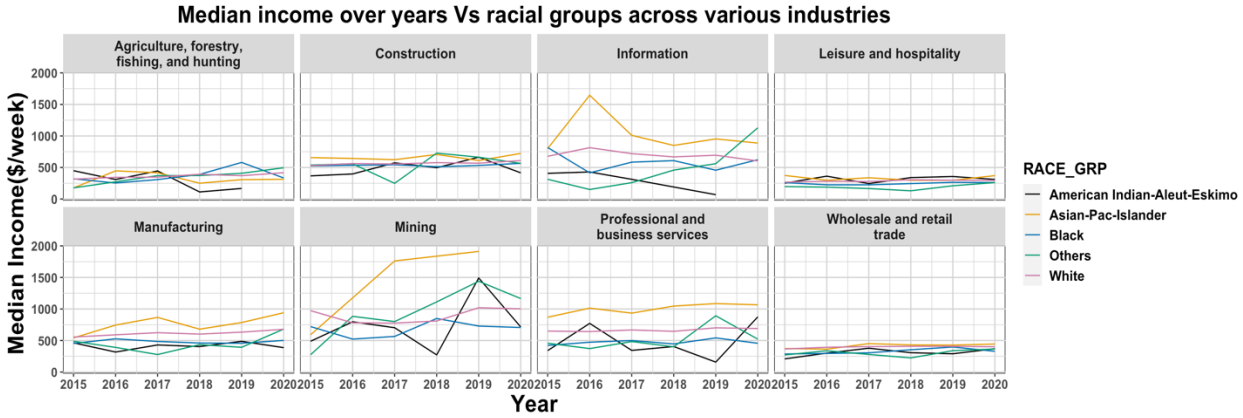


Figure 7

Change in Median Income by Age over years

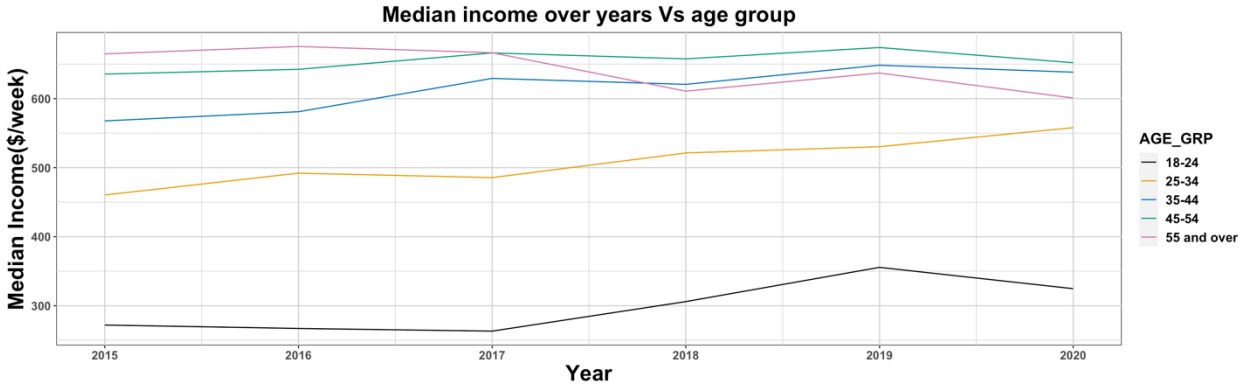


Figure 8

In the figure 8, we observe that with age, the median income rises. In the 2019-2020 year, the age group 25-34 has witnessed an increase in median income, whereas the rest of the categories have seen decreases, during this same period, the age groups 18-24 and 55 and over showed a considerable reduction, implying that most young and older employees with higher income might have lost their jobs during their epidemic. Under 18 group is not considered in our analysis due to missing income data in few industries.

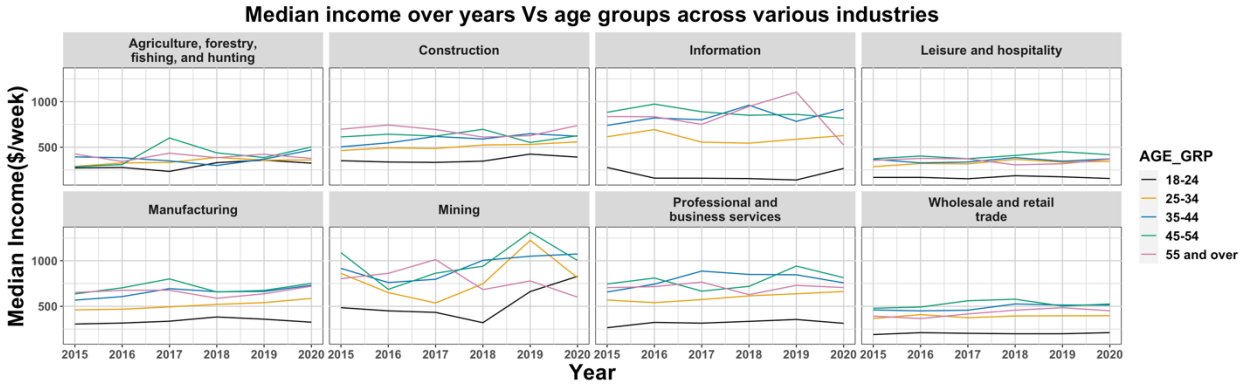


Figure 9

In contrast to the general trend, the plot in figure 9 show that the 18-24 age group experienced a big salary gain in the mining industry and a slight increase in the information business, while they were unaffected in the leisure and hospitality and wholesale and retail trade industries. Wages in the information and mining industries plummeted for those aged 55 and over, while they fared better in manufacturing, construction, and leisure and hospitality.

Change in Median Income by Educational attainment over years

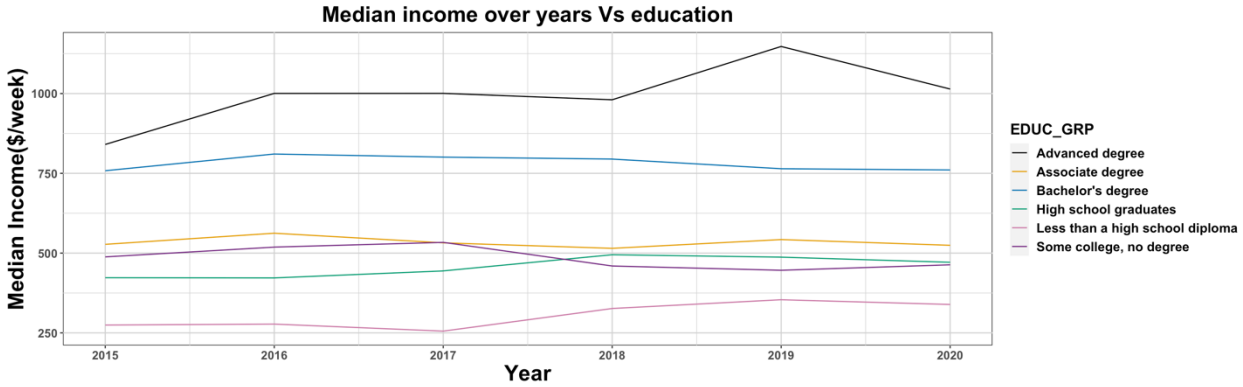


Figure 10

In general, highly qualified people earn more money than the rest of the population. People with advanced degrees, such as a master's or Advanced degree, earn more than those with only a bachelor's degree. People with a high school diploma or less earn the least, and people with an Associate degree earn more money, but not nearly as much as those with a bachelor's degree or another advanced degree. Overall, the median Income for high school graduates shows an increasing trend between 2015 and 2020.

Does having a higher qualification helps earn income in every industry?

We see in figure 11 that people with greater qualifications earn more in general across all industries. People with greater qualifications are likely to earn much more in manufacturing, information, mining, professional and business services, and wholesale and retail trade than people with lower qualifications, whereas the difference is not as large in other industries. In every field, persons with less than a high school graduation earn the least.

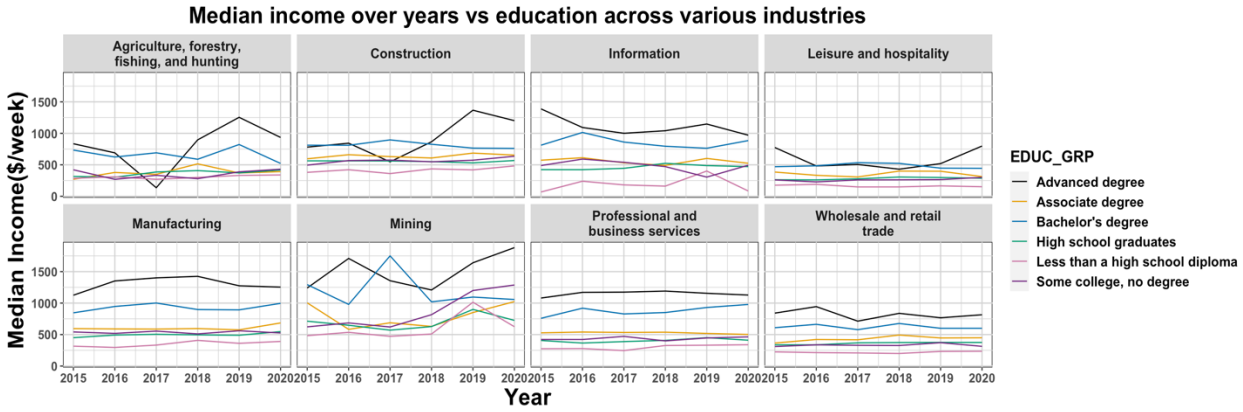


Figure 11

Change in Median Income over years based on gender and racial groups

Do females across all races earn lesser their male counterparts?

For both genders, Asian-Pacific Islanders earn more than other racial groupings. Males of a certain racial group make more than females in general; Asian-Pacific-Islander females earn more than males of the other four racial groups.

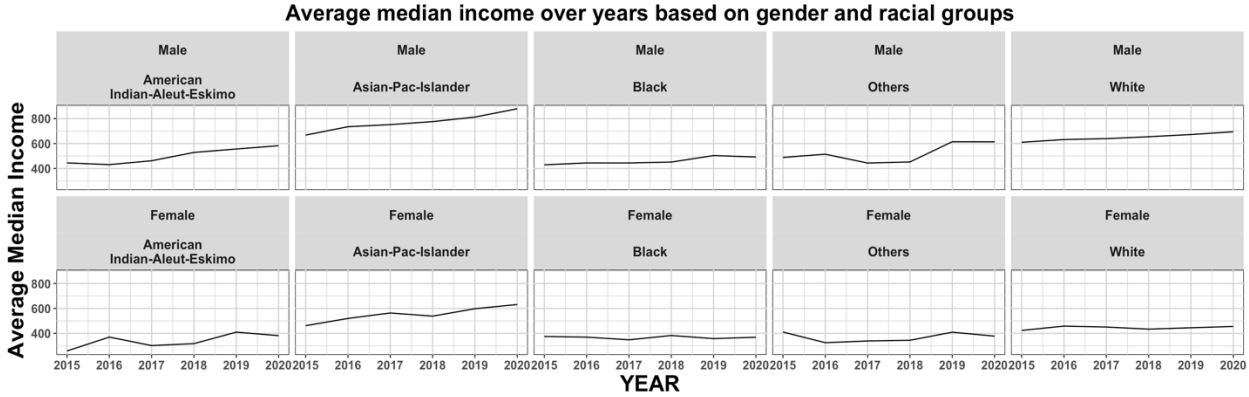


Figure 12

2. Fitting a model to predict the median income based on these demographic variables.

Choice of model

A linear relationship between an explanatory variable (X) and a response variable is assumed in a linear regression model (y). Based on the value of X, we can anticipate the value of y. In our context here, our response variable y is the median income, while the explanatory variables are year, gender, age group, major industry, race group, and education group. Using a RLM model did not improved our model and hence we decided to stick with a linear model.

Process

We found that the median income is right skewed which can be in Appendix Figure (a). We fit a linear model to revenue after applying a log transformation to make it normally distributed (Appendix Figure (b)). With log converted income data, we found a considerable improvement in the model.

Initially, we had fit a simple model that had all the 5 demographic variables without any interactions. The residual plot of this model in Appendix Figure (c) shows a straight line going through the x-axis, with a small increasing trend towards the right, and the AIC value is 30511.6. The residuals are roughly evenly

distributed as well, however the bottom right half of the plot has a few more points than the top right half. By introducing interactions between the predictors, we hope to improve the model.

3. Verify whether introducing interactions between the demographic variables improve our model.

Check if Major Industry and Gender interact

We looked at how Gender and Major Industry interact from Figure 13 and discovered that they had a substantial relationship. We've seen how other variables interact with one another in the same way which can be observed from Appendix Plot (d to f).

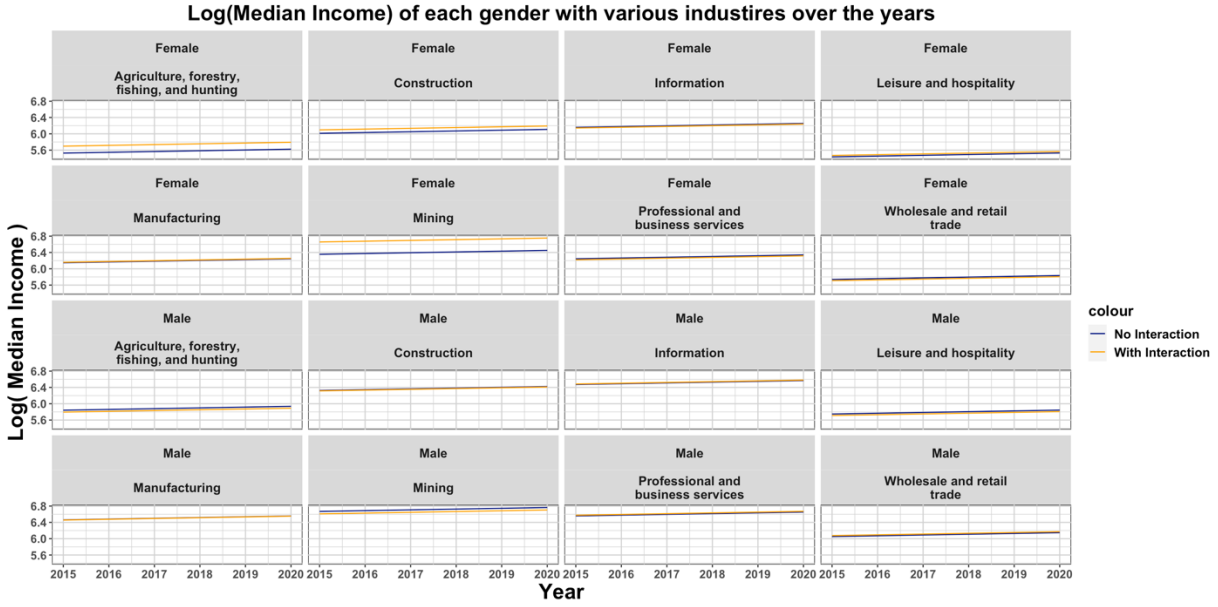


Figure 13

We saw major industries do interact with gender, age group, race group, education level and in our analysis, we also noticed interactions between age group with race group and gender. By adding these meaningful interactions to our model, we saw the AIC value reduced to 24114.14 (summary of the model is available in the appendix), which is much better than the model without interaction having an AIC value of 30511.6. Comparing with the previous model's residual plots, the residual plot (Appendix Figure g) of this model with interaction leads to a smoother with a straight line about the x-axis with no bends, and the residuals are distributed evenly, suggesting that our model fits the data well, and is the model we have chosen to predict the median income. The figure 14 below shows the Log (Median Income) as a response variable for each racial group and gender over the years.

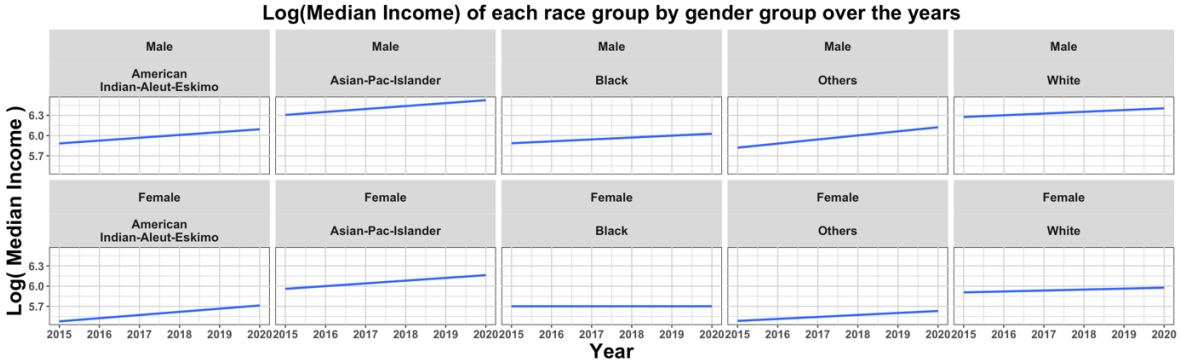


Figure 14

We experimented adding 3-way interactions to verify if our improves. The AIC value does decrease however, from the residual plot in Appendix Figure (h), we can see that the model is overfitting the data which is evident from the downwards bent of the smoother in the residual plot. Hence, adding 3-way interactions over fits the model with our data and shouldn't be used.

4. Make predictions on the model.

From our analysis, we have chosen to use the model with interactions which include explanatory variables education, age, race, gender, and major industry with interactions of major industry with gender, race group, age group, education attainment, and interactions of age group with race group and gender to predict the median income.

We draw predictions for the “Agriculture, forestry, fishing, and hunting”, “Mining” and “Information’s” industries highlighting the difference in median incomes, and considering the “Black”, “White” and “Asian-Pac-Islander” race groups as they show significantly higher earnings compared to other races. Similarly, we are looking at the workforce with a high educational qualification like “Advanced Degree” and “Bachelor's Degree”. We chose the age categories "25-34", "35-44," and "45-54" of both the genders since they generate most of the country's income and are the leading labor force contributors when compared to other age groups.

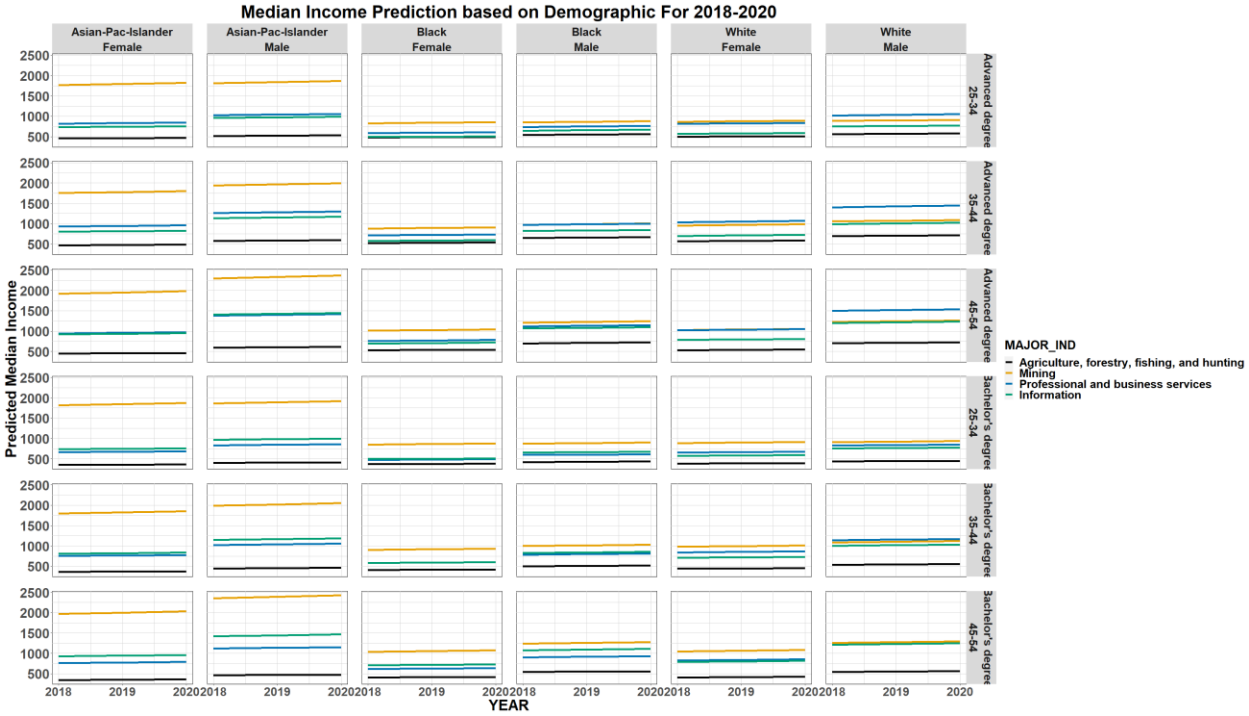


Figure 15

From Figure 15 we notice that the median income of mining has outperformed other industries, and that male Asian-Pac-Islander’s race group within age 45-54 with bachelor’s degree show the highest weekly earnings above 2400$ from 2018 to 2020, closely followed by those with Advanced degrees.

The lowest median income that is less than 500$ is earned by female Asian-Pac-Islander’s working in agriculture, forest, fishing, and mining in the age group of 45-54 with a bachelor’s degree. We see that median income earned by black females in the age group of 25-34 with an Advanced degree coincides for both industries of information and agriculture, forestry, fishing, and hunting. Further, in figure 16 the model predicts that males across each industry earn more than females, which aligns with our analysis.

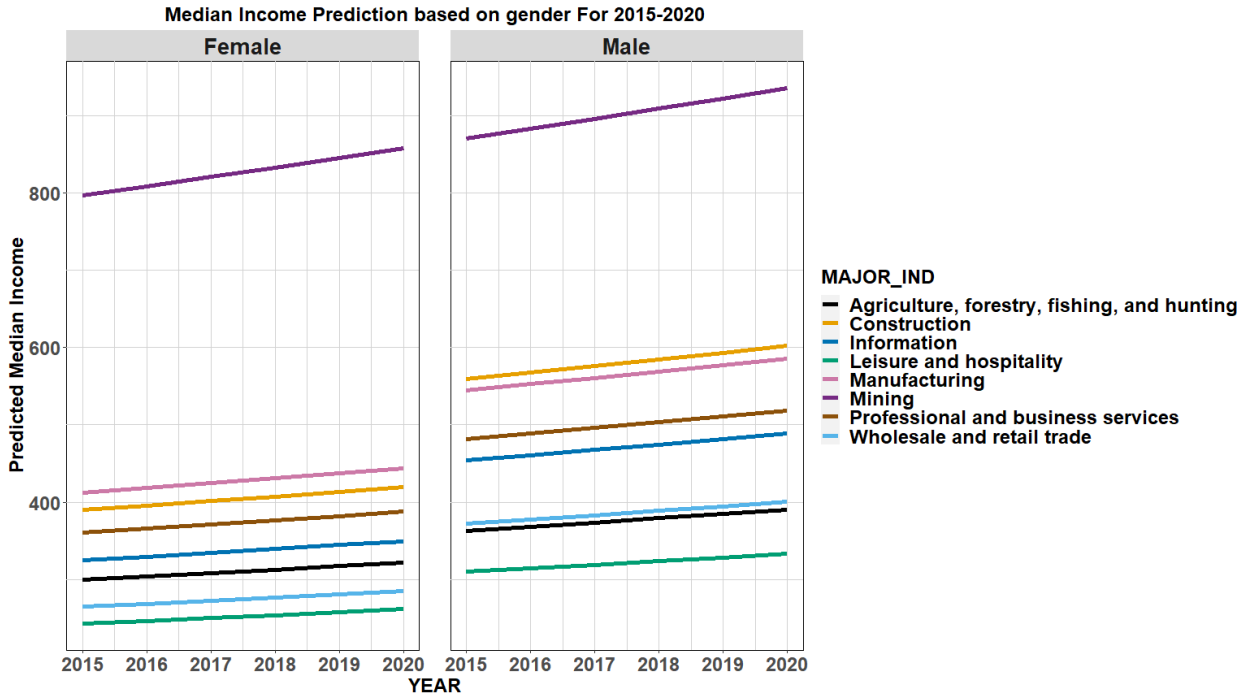


Figure 16

Conclusion and Future work

From 2015 to 2019, the unemployment rate appeared to be decreasing; but, in 2020, we saw an increase in unemployment in almost all industry sectors, except for agriculture, forestry, fishing, and hunting. The mining industry's unemployment peaked in 2016, although it appears to have recovered till 2019. Although it appears to be dropping from 2017 to 2019, the overall median income has risen in recent years, and has continued to climb during the pandemic year, implying that lower-paying jobs are more likely to be terminated during the outbreak. Over time, the mining industry has paid the most, while leisure and hospitality has paid the least. Furthermore, the median income in the mining business has fluctuated the most over the years, although most other sectors have remained consistent.

Although there is a salary disparity between men and women, this tendency does not apply to all industries. Females of the Asian-Pac-Islander racial group earn more than males of other races, whereas males of the Asian-Pac-Islander group earn less. The largest salary disparity is in the information and professional and business services industries, not in agriculture, forestry, fishing, and hunting, construction, or mining. In practically every sector, Asian-Pacific Islanders earn the most, whereas American Indian-Aleut-Eskimos earn the least, with Asian-Pacific-Islanders earning much more in mining, professional and business services, manufacturing, and information.

In all industries, median income appears to climb with age, with information and mining having the highest salary age depending on age, implying that experience is valued more in those industries. Across all industries, age groups, genders, and races, those with a higher education are more likely to earn.

Race, gender, age, and educational attainment all play a substantial impact in determining median income across all industries and may be well predicted using a linear model rather of more complex modeling such as loess. Interactions between major industries and gender, race, age, and educational attainment, as well as interactions between age groups and race and gender, improve model regression.

The fitted model's predictions are consistent with the analysis, indicating that a salary difference exists between men and women. Furthermore, Asian-Pacific Islanders, both male and female, earn much more in the mining business than other racial groupings in other industries.

However, a more thorough investigation is needed to comprehend the wage disparity based on demographics, considering factors such as job description, location of employment, and experience in the field, to provide a more detailed analysis to the current underrepresented groups in the labor force.

References

1) https://drive.google.com/file/d/1R5_A_vr47x8Nw55-c7YbrAy3IJO-qolK/view
2) https://assets.ipums.org/_files/exercises/ipums-usa-exercise-r-1.pdf
3) https://cps.ipums.org/cps/cpi99.shtml
4) https://www.bls.gov/charts/employment-situation/civilian-unemployment-rate.htm

Appendix

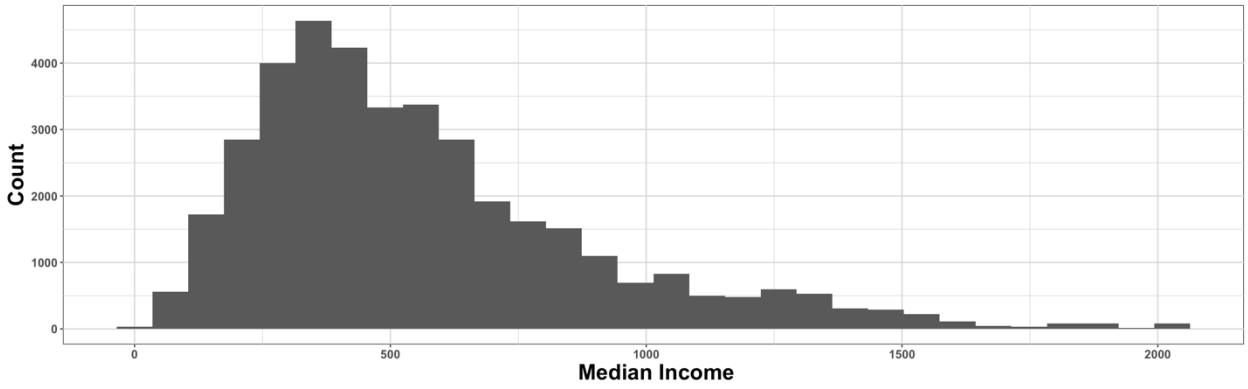


Figure (a)

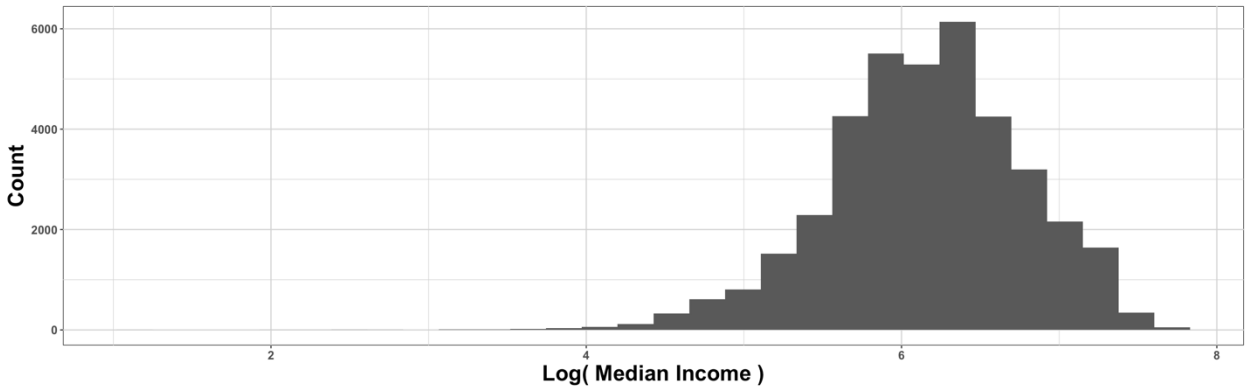


Figure (b)

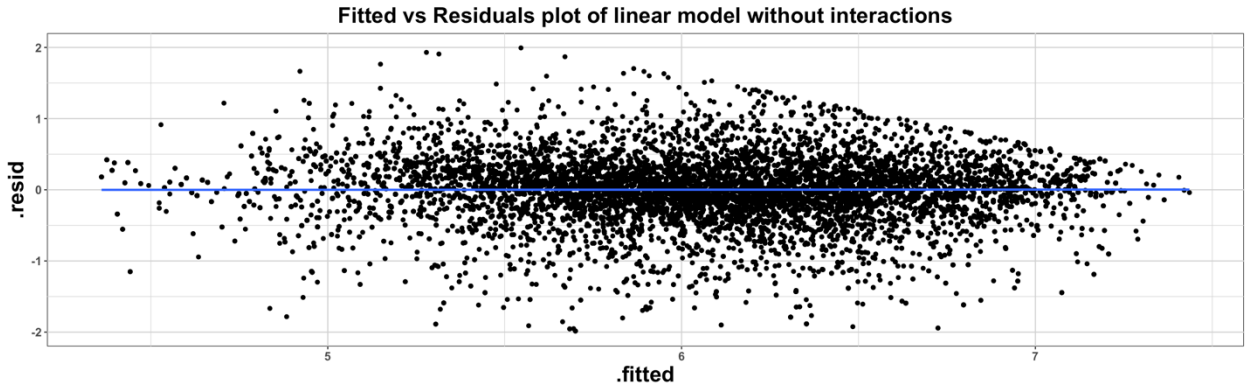


Figure (c) (Linear Model without interactions)

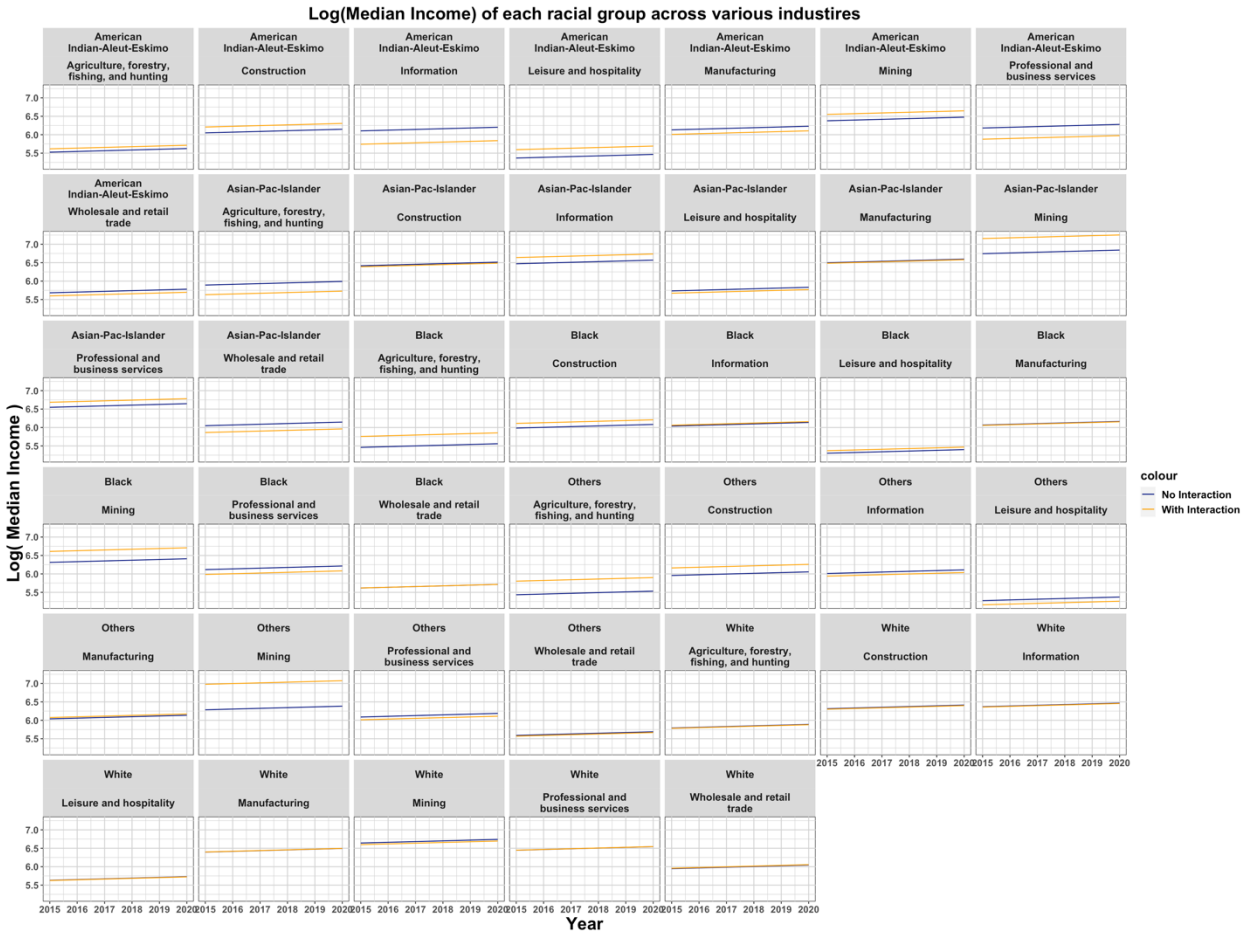
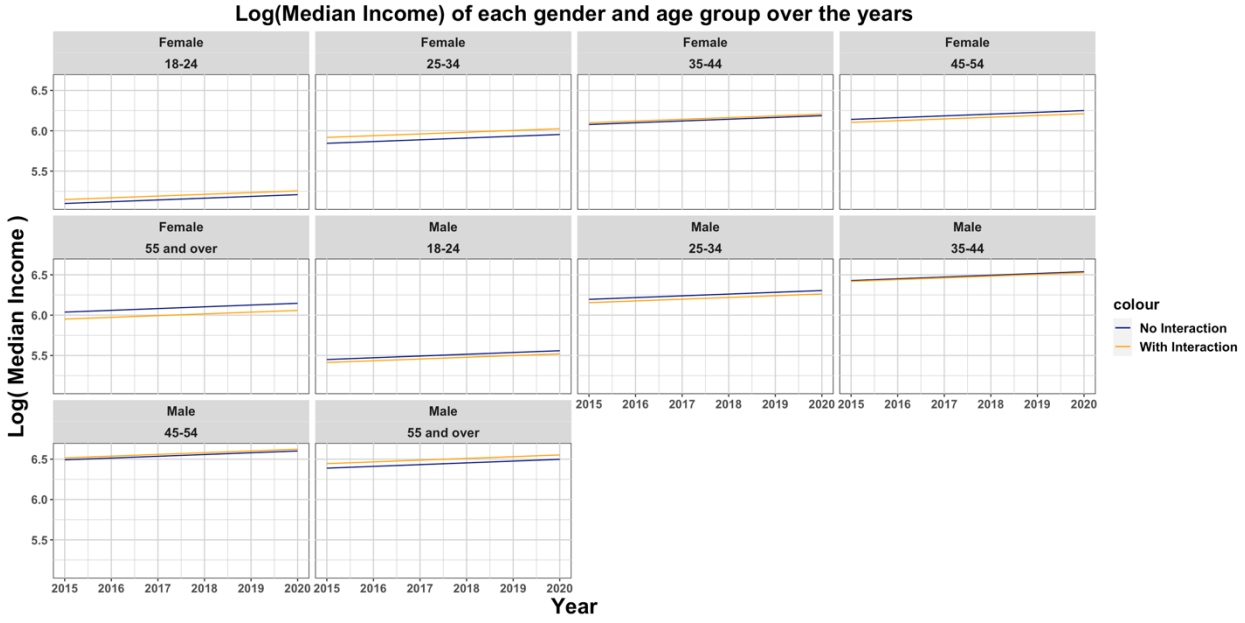


Figure (d)



Figure(e)

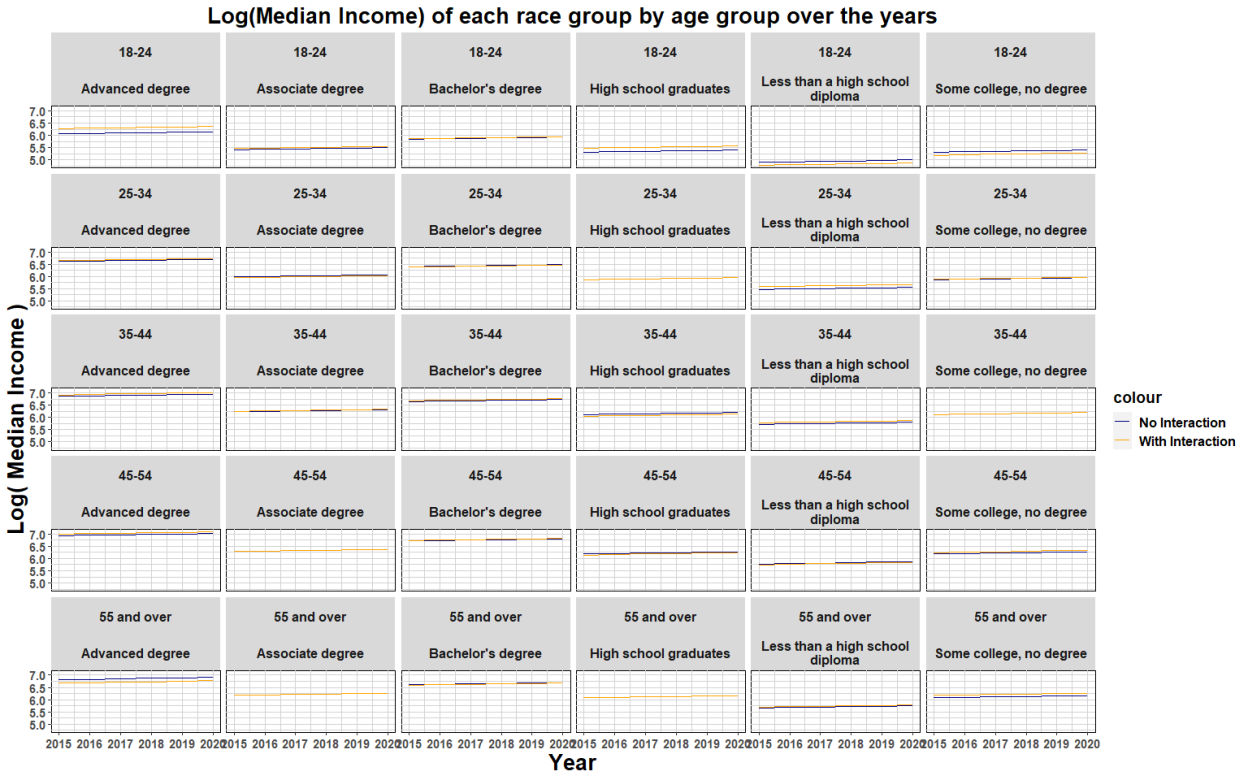


Figure (f)

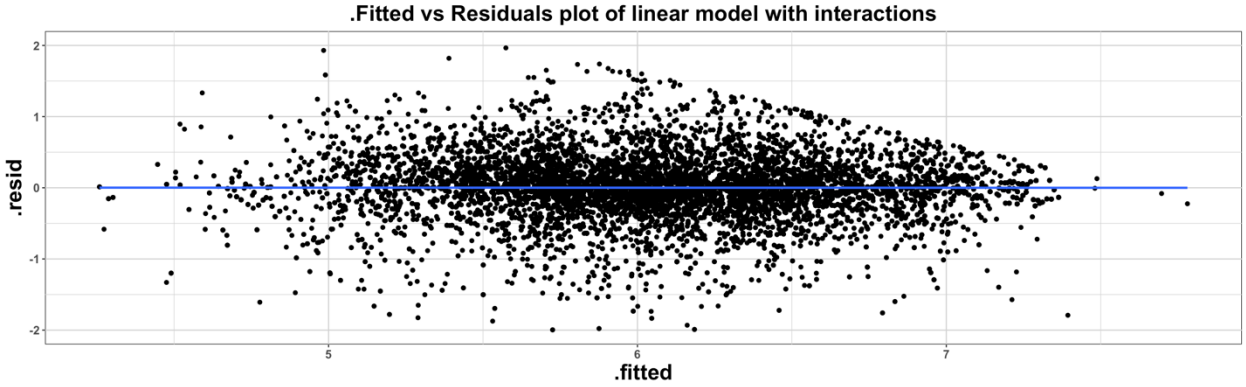


Figure (g)

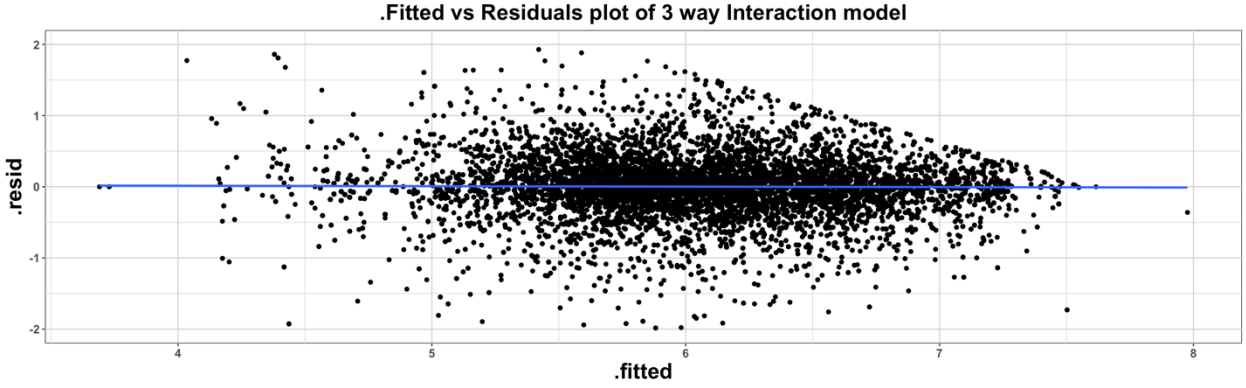


Figure (h)

Summary of the linear model with interactions

Weighted Residuals:

Min	1Q	Median	3Q	Max
-222.306	-4.699	0.009	5.177	147.695

Coefficients:

	Estimate	Std.Error	t value	t Value
(Intercept)	-23.215494	1.775831	-13.073	< 2e-16
GENDERFemale	-0.074045	0.029759	-2.488	0.012844
YEAR	0.014566	0.000879	16.571	< 2e-16
RACE_GRPAsian-Pac-Islander	-0.315644	0.116785	-2.703	0.006879
RACE_GRPBlack	0.013444	0.111719	0.12	0.904215
RACE_GRPOthers	0.025342	0.145266	0.174	0.861509
RACE_GRPWhite	-0.14643	0.095362	-1.536	0.124664
EDUC_GRPAssociate degree	-0.517383	0.078623	-6.581	4.75E-11
EDUC_GRPBachelor's degree	-0.257121	0.070561	-3.644	0.000269
EDUC_GRPHigh school graduates	-0.509046	0.066407	-7.666	1.82E-14
EDUC_GRPLess than a high school diploma	-0.813621	0.068017	-11.962	< 2e-16
EDUC_GRPSome college	-0.617866	0.072019	-8.579	< 2e-16
MAJOR_INDConstruction	0.301749	0.125903	2.397	0.016549
MAJOR_INDInformation	-0.458776	0.150225	-3.054	0.00226
MAJOR_INDLeisure and hospitality	-0.421391	0.121471	-3.469	0.000523
MAJOR_INDManufacturing	0.458188	0.122666	3.735	0.000188
MAJOR_INDMining	0.5181	0.169026	3.065	0.002177
MAJOR_INDProfessional and business services	0.269537	0.125848	2.142	0.032218
MAJOR_INDWholesale and retail trade	-0.462813	0.12097	-3.826	0.000131
AGE_GRP25-34	-0.053588	0.063798	-0.84	0.40093
AGE_GRP35-44	0.459231	0.06499	7.066	1.62E-12
AGE_GRP45-54	0.431543	0.068738	6.278	3.46E-10
AGE_GRP55 and over	0.298416	0.070876	4.21	2.55E-05
GENDERFemale:MAJOR_INDConstruction	-0.169191	0.032078	-5.274	1.34E-07
GENDERFemale:MAJOR_INDInformation	-0.142785	0.03225	-4.427	9.56E-06
GENDERFemale:MAJOR_INDLeisure and hospitality	-0.050863	0.029457	-1.727	0.08423
GENDERFemale:MAJOR_INDManufacturing	-0.085594	0.029396	-2.912	0.003596
GENDERFemale:MAJOR_INDMining	0.104252	0.049682	2.098	0.035879
GENDERFemale:MAJOR_INDProfessional and business services	-0.098052	0.029204	-3.358	0.000787
GENDERFemale:MAJOR_INDWholesale and retail trade	-0.147563	0.02905	-5.08	3.80E-07

RACE_GRPAsian-Pac-Islander:MAJOR_INDConstruction	0.157855	0.12114	1.303	0.192555
RACE_GRPBlack:MAJOR_INDConstruction	-0.179145	0.113924	-1.573	0.115842
RACE_GRPOthers:MAJOR_INDConstruction	-0.14719	0.148364	-0.992	0.321163
RACE_GRPWhite:MAJOR_INDConstruction	-0.002963	0.096987	-0.031	0.975629
RACE_GRPAsian-Pac-Islander:MAJOR_INDInformation	0.459508	0.144972	3.17	0.001527
RACE_GRPBlack:MAJOR_INDInformation	0.019432	0.140582	0.138	0.890064
RACE_GRPOthers:MAJOR_INDInformation	-0.026272	0.17668	-0.149	0.881795
RACE_GRPWhite:MAJOR_INDInformation	0.131615	0.126788	1.038	0.299241
RACE_GRPAsian-Pac-Islander:MAJOR_INDLeisure and hospitality	-0.016668	0.114659	-0.145	0.884421
RACE_GRPBlack:MAJOR_INDLeisure and hospitality	-0.289741	0.110066	-2.632	0.008481
RACE_GRPOthers:MAJOR_INDLeisure and hospitality	-0.469222	0.144771	-3.241	0.001192
RACE_GRPWhite:MAJOR_INDLeisure and hospitality	-0.070961	0.093726	-0.757	0.448987
RACE_GRPAsian-Pac-Islander:MAJOR_INDManufacturing	0.176303	0.116241	1.517	0.129351
RACE_GRPBlack:MAJOR_INDManufacturing	-0.064311	0.111963	-0.574	0.565704
RACE_GRPOthers:MAJOR_INDManufacturing	-0.070901	0.147334	-0.481	0.63036
RACE_GRPWhite:MAJOR_INDManufacturing	0.137666	0.095922	1.435	0.15124
RACE_GRPAsian-Pac-Islander:MAJOR_INDMining	0.611717	0.197062	3.104	0.001909
RACE_GRPBlack:MAJOR_INDMining	-0.192892	0.167573	-1.151	0.2497
RACE_GRPOthers:MAJOR_INDMining	0.089072	0.20359	0.438	0.661747
RACE_GRPWhite:MAJOR_INDMining	-0.186427	0.143174	-1.302	0.192888
RACE_GRPAsian-Pac-Islander:MAJOR_INDProfessional and business services	0.309524	0.119445	2.591	0.009564
RACE_GRPBlack:MAJOR_INDProfessional and business services	-0.069635	0.115445	-0.603	0.546386
RACE_GRPOthers:MAJOR_INDProfessional and business services	-0.078869	0.149565	-0.527	0.59797
RACE_GRPWhite:MAJOR_INDProfessional and business services	0.223124	0.100065	2.23	0.025767
RACE_GRPAsian-Pac-Islander:MAJOR_INDWholesale and retail trade	0.197592	0.114573	1.725	0.084611
RACE_GRPBlack:MAJOR_INDWholesale and retail trade	-0.01279	0.110091	-0.116	0.907517
RACE_GRPOthers:MAJOR_INDWholesale and retail trade	-0.078236	0.143968	-0.543	0.586838
RACE_GRPWhite:MAJOR_INDWholesale and retail trade	0.23981	0.09393	2.553	0.010681
MAJOR_INDConstruction:AGE_GRP25-34	0.049659	0.046503	1.068	0.285585
MAJOR_INDInformation:AGE_GRP25-34	0.612809	0.052735	11.621	< 2e-16
MAJOR_INDLeisure and hospitality:AGE_GRP25-34	0.300505	0.043979	6.833	8.44E-12
MAJOR_INDManufacturing:AGE_GRP25-34	0.012137	0.045228	0.268	0.788432
MAJOR_INDMining:AGE_GRP25-34	0.12221	0.079608	1.535	0.124756
MAJOR_INDProfessional and business services:AGE_GRP25-34	0.10389	0.044969	2.31	0.020879
MAJOR_INDWholesale and retail trade:AGE_GRP25-34	0.312326	0.043814	7.128	1.03E-12
MAJOR_INDConstruction:AGE_GRP35-44	-0.011204	0.048006	-0.233	0.815463
MAJOR_INDInformation:AGE_GRP35-44	0.678505	0.054414	12.469	< 2e-16
MAJOR_INDLeisure and hospitality:AGE_GRP35-44	0.242	0.046066	5.253	1.50E-07

MAJOR_INDManufacturing:AGE_GRP35-44	0.067468	0.046826	1.441	0.14964
MAJOR_INDMining:AGE_GRP35-44	0.084174	0.080755	1.042	0.297261
MAJOR_INDProfessional and business services:AGE_GRP35-44	0.205264	0.046708	4.395	1.11E-05
MAJOR_INDWholesale and retail trade:AGE_GRP35-44	0.398623	0.04566	8.73	< 2e-16
MAJOR_INDConstruction:AGE_GRP45-54	0.053157	0.047668	1.115	0.26479
MAJOR_INDInformation:AGE_GRP45-54	0.862202	0.054031	15.957	< 2e-16
MAJOR_INDLeisure and hospitality:AGE_GRP45-54	0.392083	0.045706	8.578	< 2e-16
MAJOR_INDManufacturing:AGE_GRP45-54	0.18506	0.046143	4.011	6.07E-05
MAJOR_INDMining:AGE_GRP45-54	0.22428	0.081482	2.753	0.005917
MAJOR_INDProfessional and business services:AGE_GRP45-54	0.264793	0.046199	5.732	1.00E-08
MAJOR_INDWholesale and retail trade:AGE_GRP45-54	0.443959	0.045004	9.865	< 2e-16
MAJOR_INDConstruction:AGE_GRP55 and over	0.145685	0.046345	3.143	0.001671
MAJOR_INDInformation:AGE_GRP55 and over	0.777729	0.052947	14.689	< 2e-16
MAJOR_INDLeisure and hospitality:AGE_GRP55 and over	0.140415	0.044475	3.157	0.001594
MAJOR_INDManufacturing:AGE_GRP55 and over	0.153045	0.044909	3.408	0.000655
MAJOR_INDMining:AGE_GRP55 and over	0.110978	0.07975	1.392	0.164059
MAJOR_INDProfessional and business services:AGE_GRP55 and over	0.202175	0.044911	4.502	6.76E-06
MAJOR_INDWholesale and retail trade:AGE_GRP55 and over	0.255571	0.043664	5.853	4.86E-09
EDUC_GRPAssociate degree:MAJOR_INDConstruction	0.162983	0.084091	1.938	0.052609
EDUC_GRPBachelor's degree:MAJOR_INDConstruction	0.208131	0.076145	2.733	0.006272
EDUC_GRPHigh school graduates:MAJOR_INDConstruction	0.076386	0.071659	1.066	0.286441
EDUC_GRPLess than a high school diploma:MAJOR_INDConstruction	0.020913	0.073914	0.283	0.777223
EDUC_GRPAssociate degree:MAJOR_INDInformation	-0.019397	0.084974	-0.228	0.81944
EDUC_GRPBachelor's degree:MAJOR_INDInformation	0.266831	0.074114	3.6	0.000318
EDUC_GRPHigh school graduates:MAJOR_INDInformation	-0.077125	0.072247	-1.068	0.285742
EDUC_GRPLess than a high school diploma:MAJOR_INDInformation	-0.371764	0.085539	-4.346	1.39E-05
EDUC_GRPAssociate degree:MAJOR_INDLeisure and hospitality	0.300975	0.08192	3.674	0.000239
EDUC_GRPBachelor's degree:MAJOR_INDLeisure and hospitality	0.330203	0.073515	4.492	7.09E-06
EDUC_GRPHigh school graduates:MAJOR_INDLeisure and hospitality	0.226886	0.069318	3.273	0.001065
EDUC_GRPLess than a high school diploma:MAJOR_INDLeisure and hospitality	0.116141	0.071377	1.627	0.103713
EDUC_GRPAssociate degree:MAJOR_INDManufacturing	-0.186453	0.08002	-2.33	0.019807
EDUC_GRPBachelor's degree:MAJOR_INDManufacturing	0.009332	0.071766	0.13	0.896545
EDUC_GRPHigh school graduates:MAJOR_INDManufacturing	-0.328664	0.067567	-4.864	1.15E-06
EDUC_GRPLess than a high school diploma:MAJOR_INDManufacturing	-0.398072	0.070179	-5.672	1.42E-08
EDUC_GRPAssociate degree:MAJOR_INDMining	0.246544	0.098945	2.492	0.012716
EDUC_GRPBachelor's degree:MAJOR_INDMining	0.282529	0.089173	3.168	0.001534
EDUC_GRPHigh school graduates:MAJOR_INDMining	0.154853	0.084152	1.84	0.065752
EDUC_GRPLess than a high school diploma:MAJOR_INDMining	0.125987	0.101033	1.247	0.212412

EDUC_GRPAssociate degree:MAJOR_INDProfessional and business services	-0.141081	0.079756	-1.769	0.076918
EDUC_GRPBachelor's degree:MAJOR_INDProfessional and business services	0.046006	0.071178	0.646	0.518055
EDUC_GRPHigh school graduates:MAJOR_INDProfessional and business services	-0.398467	0.067339	-5.917	3.30E-09
EDUC_GRPLess than a high school diploma:MAJOR_INDProfessional and business services	-0.628117	0.070545	-8.904	< 2e-16
EDUC_GRPAssociate degree:MAJOR_INDWholesale and retail trade	0.165349	0.080302	2.059	0.039493
EDUC_GRPBachelor's degree:MAJOR_INDWholesale and retail trade	0.289402	0.072168	4.01	6.08E-05
EDUC_GRPHigh school graduates:MAJOR_INDWholesale and retail trade	0.11639	0.067997	1.712	0.086961
EDUC_GRPLess than a high school diploma:MAJOR_INDWholesale and retail trade	0.029144	0.070414	0.414	0.678956
RACE_GRPAsian-Pac-Islander:AGE_GRP25-34	0.445137	0.054772	8.127	4.53E-16
RACE_GRPBlack:AGE_GRP25-34	0.162528	0.050731	3.204	0.001358
RACE_GRPOthers:AGE_GRP25-34	0.085579	0.058537	1.462	0.143759
RACE_GRPWhite:AGE_GRP25-34	0.356325	0.048548	7.34	2.18E-13
RACE_GRPAsian-Pac-Islander:AGE_GRP35-44	0.035823	0.055066	0.651	0.515347
RACE_GRPBlack:AGE_GRP35-44	-0.177629	0.051112	-3.475	0.000511
RACE_GRPOthers:AGE_GRP35-44	0.052387	0.060861	0.861	0.389376
RACE_GRPWhite:AGE_GRP35-44	0.058832	0.04855	1.212	0.225597
RACE_GRPAsian-Pac-Islander:AGE_GRP45-54	0.092692	0.059848	1.549	0.121438
RACE_GRPBlack:AGE_GRP45-54	-0.072438	0.056265	-1.287	0.197944
RACE_GRPOthers:AGE_GRP45-54	0.12509	0.068301	1.831	0.067041
RACE_GRPWhite:AGE_GRP45-54	0.090428	0.053845	1.679	0.093079
RACE_GRPAsian-Pac-Islander:AGE_GRP55 and over	0.121553	0.06363	1.91	0.056098
RACE_GRPBlack:AGE_GRP55 and over	0.050757	0.059981	0.846	0.397433
RACE_GRPOthers:AGE_GRP55 and over	0.191953	0.07406	2.592	0.009549
RACE_GRPWhite:AGE_GRP55 and over	0.194136	0.057406	3.382	0.000721
GENDERFemale:AGE_GRP25-34	-0.055558	0.01086	-5.116	3.14E-07
GENDERFemale:AGE_GRP35-44	-0.131352	0.011714	-11.213	< 2e-16
GENDERFemale:AGE_GRP45-54	-0.208625	0.011578	-18.019	< 2e-16
GENDERFemale:AGE_GRP55 and over	-0.19269	0.011129	-17.314	< 2e-16

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13.08 on 38518 degrees of freedom

Multiple R-squared: 0.7712, Adjusted R-squared: 0.7703

F-statistic: 927.2 on 140 and 38518 DF, p-value: < 2.2e-16