# Consultancy Project

Final report

Team 7 GlassWage: Be Rich!

Zhizun (Skye) Li, Linan (Iris) Zhou

1. Introduction	2
1.1 Field of Consultancy	2
1.2 Potential Clients	2
1.3 Data Sources	2
2. Database Design	3
2.1 Tables	3
2.2 EER Diagram	11
3. Data Manipulation and Visualization	12
3.1 Cross-sectional Analysis	12
3.1.1 Geographical distribution	12
3.1.2 Top ten metropolitan areas	13
3.1.3 Wage distribution among selected job titles	14
3.1.4 Wage growth rate among job levels	15
3.2 Time Series Analysis	16
3.2.1 Dimension 1: the level of the occupation(WageLevel)	17
3.2.2 Dimension 2: the geographical area(Area)	19
3.2.3 Dimension 3: the title of the occupation(SocCode)	22
4. Summary	25

## 1. Introduction

## 1.1 Field of Consultancy

This project aims to help job seekers obtain a reasonable expectation of salary package when hunting for jobs in the United States. By referring to our database, job seekers will be able to improve their bargaining power in the labor market.

Our project focuses on the prevailing wages of data analytics positions, covering domains such as financial analytics, marketing analytics and database management. Our job selection considerations are based on the career outcomes of previous IBS graduates.

Our database includes the prevailing wages of data analytics positions from 7/2017 to 6/2021. By analyzing time-series data, we can see the overall growth trend of wages in data analytics positions and identify data positions that have the greatest growth potential.

The database also provides job requirements of job positions included. We can help students who are interested in data-related positions better prepare for graduate recruitment season.

#### 1.2 Potential Clients

Our potential clients are job seekers and students who want to land full-time data analytics jobs in the United States. Employers who are interested in data-related recruitment can also refer to our database for salary budget surveys.

#### 1.3 Data Sources

The prevailing wages in this database originate from <u>Foreign Labor Certification Data Center</u>. The prevailing wages are gathered and calculated by the Office of Foreign Labor Certification, Employment and Training Administration of the U.S. Department of Labor. They use the prevailing wages as an indicator of reasonable salary paid to employees when they issue certain visas like H1b to foreign employees in the U.S. Sources of original data and detailed explanation can be found in the <u>U.S.</u> <u>Department of Labor</u>.

This database uses Standard Occupational Classification (SOC) System and Occupational Information Network (O\*NET) to relate job titles to specific job codes, classify job titles to certain job zones. See detailed information on the U.S. Bureau of Labor Statistics and O\*NET Online.

To better reflect the actual purchasing power of wages in different periods, we added a table called "Inflation\_rate" into our database and linked it with the "pw" table using the key "Year". The original data of this table can be found on the website of IMF Data.

## 2. Database Design

Our database includes eight tables, which are "pw", "socjd", "onetjd", "jobzone", "wagelevel", "area", "state" and "inflation\_rate", connected by six relations. Please see detailed information below.

## 2.1 Tables

Table Name	pw	
Structure	158748 rows * 5 columns	
Rows	Each row represents a record of the prevailing wage of a certain job position by applicants' job level in a certain year and area	
Columns	Area: location of listed jobs	
	SocCode: job codes under SOC System	
	Year: year of prevailing wages apply to	
	WageLevel: qualification level of job applicants	
	WageRate: per hour wage in US Dollars	
Primary key	Area, SocCode, Year, WageLevel	
Foreign key	Area, SocCode, WageLevel	
Normalization	1NF check: all fields include only one single piece of data	
	2NF check: all four primary keys are needed to determine a certain WageRate	
	3NF check: only one non-key field	

#### Head of table *pw*:

Area	SocCode	Year	WageLevel	WageRate
100001	13-1111	2021	Level1	34.54
100001	13-1111	2021	Level2	37.51
100001	13-1111	2021	Level3	40.49
100001	13-1111	2021	Level4	43.46
100001	13-1161	2021	Level1	12.83

Table Name	socjd	
Structure	18 rows * 4 columns	
Rows	Each row refers a SocCode to a certain job title, description and an OnetCode	
Columns	SocCode: job codes under SOC system	
	SocTitle: job titles that SOC codes refer to	
	SocDefinition: job descriptions of job titles under SOC system	
	OnetCode: job codes under O*NET program that SOC job codes refer to	
Primary key	SocCode	
Foreign key	OnetCode	
Normalization	1NF check: one single piece of data in all fields	
	2NF check: only one primary key	
	3NF check: all non-key fields (SocTitle, SocDefinition, OnetCode) are independent of any other non-key field	

### Head of table *socjd*:

SocCode	SocTitle	SocDefinition	OnetCode
13-1111	Management Analysts	Conduct organizational studies and evaluations, design systems and procedures, conduct work simplification and measurement studies, and prepare operations and procedures manuals to assist management in operating more efficiently and effectively. Includes program analysts and management consultants.	13-1111.00
13-1161	Market Research Analysts and Marketing Specialists	Research market conditions in local, regional, or national areas, or gather information to determine potential sales of a product or service, or create a marketing campaign. May gather information on competitors, prices, sales, and methods of marketing and distribution.	13-1161.00

Table Name	onetjd
Structure	18 rows * 3 columns
Rows	Each row refers an OnetCode to a certain job title, description and job zone
Columns	OnetCode: job codes under the O*NET program
	OnetTitle: job titles that Onet codes refer to
	JobZone: five job zones according to the classification of the O*NET program
Primary key	OnetCode
Foreign key	JobZone
Normalization	1NF check: one single piece of data in all fields
	2NF check: only one primary key
	3NF check: OnetTitle and JobZone are independent of each other

## Head of table *onetjd*:

OnetCode	OnetTitle	JobZone
13-1111.00	Management Analysts	5
13-1161.00	Market Research Analysts and Marketing Specialists	4
13-2011.00	Accountants and Auditors	4
13-2031.00	Budget Analysts	4
13-2041.00	Credit Analysts	4

Table Name	jobzone	
Structure	5 rows * 5 columns	
Rows	Each row illustrates the description and requirements of a certain job zone	
Columns	JobZone: five job zones according to the classification of O*NET program	
	Description: preparation requirement of each job zone	
	Experience: work experience requirement of each job zone	
	Education: education requirement of each job zone	
	JobTraining: job training requirement of each job zone	
Primary key	JobZone	
Normalization	1NF check: all fields include only one single piece of information	
	2NF check: only one primary key	
	3NF check: all non-key fields (Description, Experience, Education, JobTraining) are independent of any other non-key field	

## Head of table *jobzone*:

JobZone	Description	Experience	Education	JobTraining
1	Little or No Preparation Needed	No previous work-related skill, knowledge, or experience is needed	A high school diploma or GED certificate	A few days to a few months of training
2	Some Preparation Needed	Some previous work-related skill, knowledge, or experience, usually not needed	A high school diploma and may require some vocational training or job-related course work	A few months to one year of working with experienced employees
3	Medium Preparation Needed	Previous work-related skill, knowledge, or experience is required	Require training in vocational schools, on-the-job experience, or an associate's degree	One or two years of training involving both on-the-job experience and informal training with experienced workers

<b>Table Name</b>	wagelevel	
Structure	4 rows * 3 columns	
Rows	Each row represents the competency level of job applicants	
Columns	Level: four wage levels according to the U.S. Department of Labor	
	Type: qualification of job applicants	
	Percentile: percentile of wage level that job applicants lie in	
Primary key	Level	
Normalization	1NF check: one single piece of data in all fields	
	2NF check: only one primary key	
	3NF check: Type and Percentile are independent of each other	

## Head of table wagelevel:

Level	Туре	Percentile
Level1	Entry/Fresher	17
Level2	Qualified	34
Level3	Experienced	50
Level4	Fully Competent	67

<b>Table Name</b>	area	
Structure	533 rows * 3 columns	
Rows	Each row represents a record of an area with its detailed area information and abbreviation of the state name	
Columns	Area: area code	
	AreaName: detailed area information	
	StateAb: state name abbreviation	
Primary key	Area	
Foreign key	StateAb	
Normalization	1NF check: one single piece of data in all fields	
	2NF check: only one primary key	
	3NF check: cannot determine the StateAb with only AreaName, vice versa	

#### Head of table *area*:

Area	AreaName	StateAb
5100004	Northwest Virginia nonmetropolitan area	VA
5100001	Southwest Virginia nonmetropolitan area	VA
5100002	Southside Virginia nonmetropolitan area	VA
40060	Richmond	VA
47900	Washington-Arlington-Alexandria	VA

Table Name	state
Structure	54 rows * 2 columns
Rows	Each row represents a state in the U.S.
Columns	StateAb: state name abbreviations
	State: state name
Primary key	StateAb
Normalization	1NF check: one single piece of data in all fields
	2NF check: only one primary key
	3NF check: only one non-key column

## Head of table *state*:

StateAb	State
VA	Virginia
WA	Washington
WV	West Virginia
WI	Wisconsin
WY	Wyoming

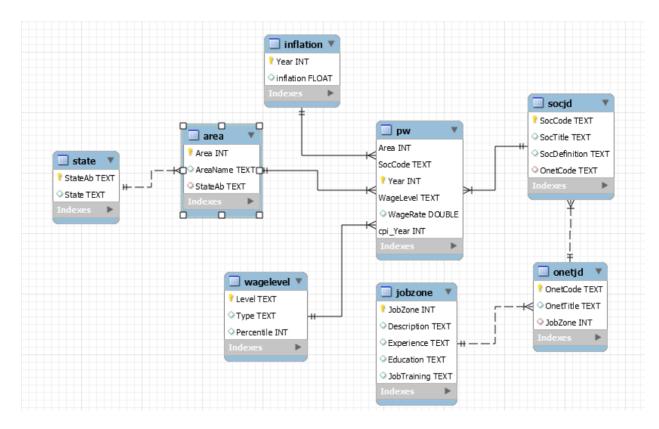
Table Name	inflation_rate	
Structure	5 rows * 2 columns	
Rows	Each row represents a record of the inflation rate in a certain year	
Columns	Year: year of prevailing wages apply to	
	inflation: the inflation rate of a certain year	
Primary key	Year	
Normalization	1NF check: all fields include only one single piece of data	
	2NF check: one primary key	
	3NF check: only one non-key field	

## The whole table of *Inflation\_rate*:

Year	Inflation
2017	0.021
2018	0.024
2019	0.018
2020	0.015
2021	0.028

## 2.2 EER Diagram

The relationship of the above-mentioned eight tables is shown below in the EER diagram.



## 3. Data Manipulation and Visualization

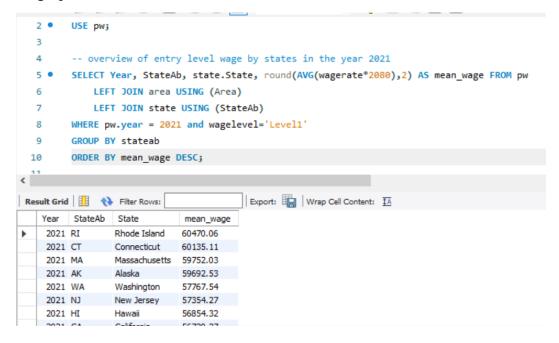
Before the actual manipulation of data, we examined if there was severe data loss in our database. Here, we used our main table "pw" as an example. Using code "pw\_pre", the result table showed that we didn't have this problem.

```
#Code: pw_pre
9 •
       SELECT *
10
       FROM pw
       WHERE Area IS NULL OR
11
12
           SocCode IS NULL OR
13
           Year IS NULL OR
           WageLevel IS NULL OR WageRate IS NULL;
14
15
esult Grid 🔢 🙌 Filter Rows:
                                        Export: Wrap Cell Content: IA
 Area SocCode Year
                      WageLevel WageRate
```

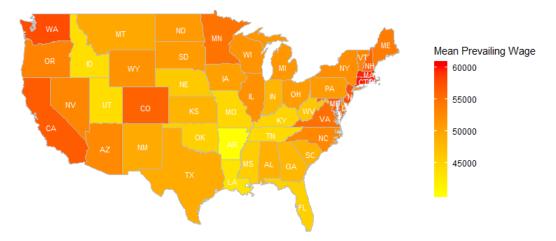
## 3.1 Cross-sectional Analysis

We selected the latest prevailing wage in 2021, conducted cross-sectional analysis, and here are some interesting findings. To show wage rates in a more straightforward way, the hourly wage rates are converted to annual wages in this part of the analysis.

#### 3.1.1 Geographical distribution

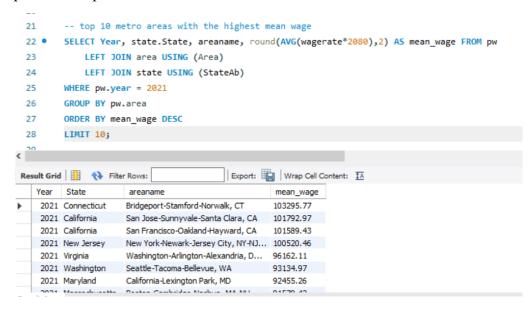


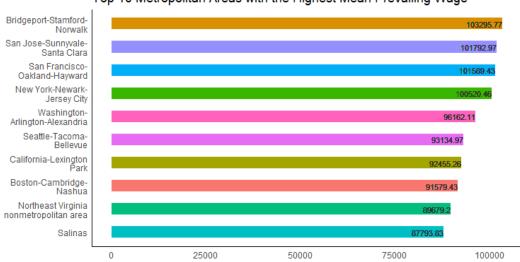
#### Mean Prevailing Wage for Entry Level Jobs by State



By manipulating data in SQL and drawing maps in Rstudio, we can plot a choropleth map of wages for entry-level business analytics jobs. The average annual wage for entry-level workers is roughly ranging from 40k-60k dollars. Most of the states with the highest wage rate are located on the western coast and eastern coast, for example, California, Washington, Massachusetts, New York and Connecticut. For job seekers, these states are also among the most popular areas when considering job locations. Besides mainstream locations, Colorado and Minnesota are also worth considering.

#### 3.1.2 Top ten metropolitan areas

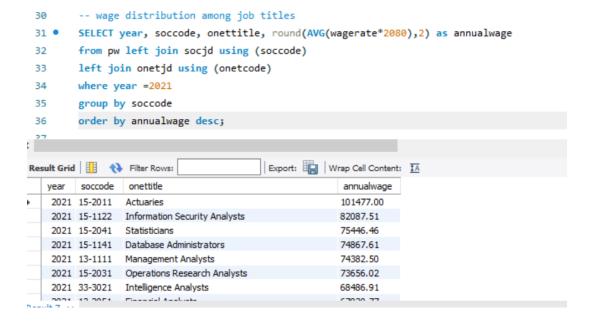


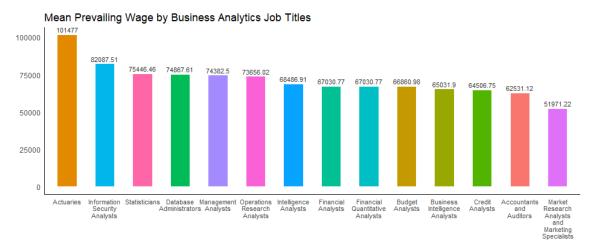


Top 10 Metropolitan Areas with the Highest Mean Prevailing Wage

Taking a step further in geographical analysis, we also selected the top ten metropolitan areas with the highest prevailing wages. Bridgeport-Stamford-Norwalk is the winning area with an annual wage of 103k dollars, and San Jose-Santa Clara and San Francisco come closely behind. It is easily seen that most highest-paid areas are located in the bay area, and areas on eastern coast such as New York and Boston. Though the competitions are extremely fierce in these areas, they are great choices for graduate students to recoup their investment in tuitions.

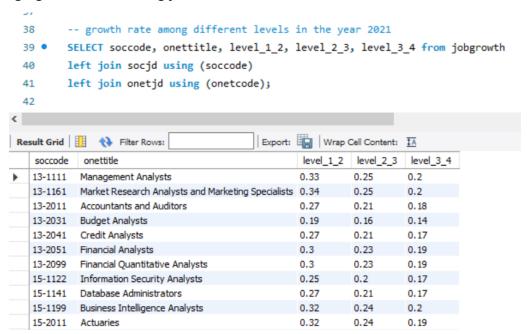
#### 3.1.3 Wage distribution among selected job titles



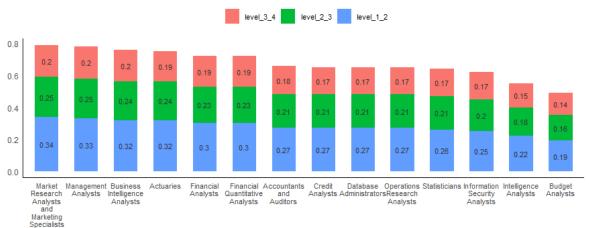


Data analytics is an interdisciplinary field that requires sufficient domain knowledge, it is powerful when it is applied to business practices. All business fields require data analysts, but the wage rates are quite different among industries. Actuaries rank the first with an obvious advantage over other jobs, followed by Information Security Analysts and Statisticians. Nevertheless, the wage rates of popular job titles such as Credit Analysts and Market Analysts are not very competitive. It can be inferred that the wage rates of jobs requiring a higher level of technical skills are higher than those that require medium-level technical skills.

#### 3.1.4 Wage growth rate among job levels







In order to research the growth potential of business analytics jobs, we calculated the growth rate among different wage levels for different job titles. On average, new graduate students are expected to increase their wage rate by 20%-30% from entry-level to senior-level positions. As the wage level increases, however, the growth rate decreases. Among job titles, Market Analysts have the highest growth rate, followed by Management Analysts and Business Intelligence Analysts. As we have mentioned in 3.1.3, the wage rates for business jobs are not competitive compared with those requiring more technical skills, but it can be seen in this figure that the growth potential is much higher. This is good news for business school graduates.

## 3.2 Time Series Analysis

Data of prevailing wages collected in our database has a period of 5 years, starting from 7/2017 and ending by 6/2021. Thus, we did a set of time series analyses to show the trends of prevailing wages and tried to describe the growth potential of wages from different dimensions.

In our main table "pw" from our database designed, we used 4 primary keys (Area, SocCode, Year, WageLevel) to determine a certain prevailing wage. In this case, we would do the majority of our time series analysis based on the following 3 dimensions: geographical area(Area), title of the occupation(SocCode) and level of the occupation(WageLevel).

To better reflect the actual purchasing power of wages in different periods, we added a table called "Inflation rate" into our database and linked it with the "pw" table using the key "Year".

#### We used the formula

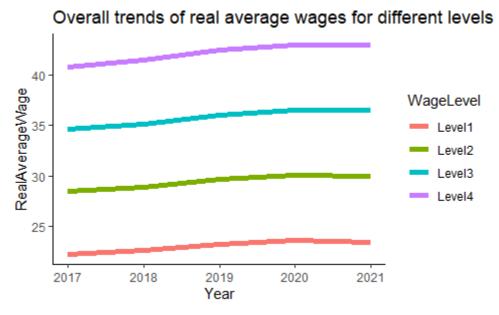
 $RealWage = (1 - Inflation) \times WageRate$ 

to calculate the real prevailing wages and got "RealAverageWage" for the former 3 dimensions. Below is an example of our SQL query.

```
# Code ts_i_1
16
17
        # For different levels of occupation
18 •
       SELECT Year, WageLevel, ROUND(AVG(WageRate) * (1 - inflation), 2) AS "RealAverageWage"
       FROM pw INNER JOIN inflation_rate AS i USING(Year)
19
20
       GROUP BY WageLevel, Year
21
       ORDER BY Year, WageLevel;
22
23
        # Code ts_i_2
24
        # For geographical areas
       SELECT Year, StateAb, ROUND(AVG(WageRate) * (1 - inflation), 2) AS "RealAverageWage"
esult Grid
             Filter Rows:
                                          Export: Wrap Cell Content: IA
                  RealAverageWage
  Year
        WageLevel
 2017
       Level1
                  22.31
 2017 Level2
 2017
       Level3
                  34.58
                  40.72
 2017
       Level1
 2018
                  22.62
 2018 Level2
                  28.88
       Level3
 2018
                  35, 15
 2018
       Laval4
```

#### 3.2.1 Dimension 1: the level of the occupation(WageLevel)

Let's start the analysis with the dimension of wage level. The line plot below shows the trends of real average wages of 4 different wage levels.

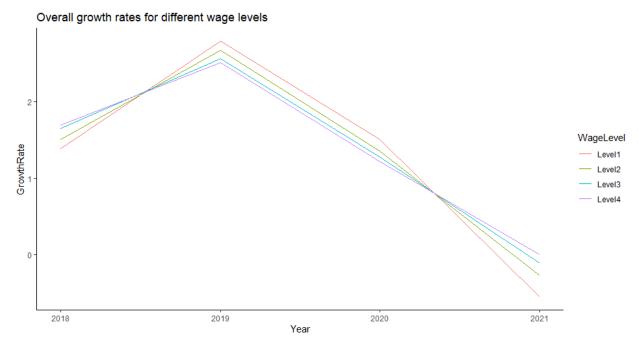


Normally, as clearly shown in this plot, higher levels gained higher wages. For all these 4 different wage levels, all real average wages increased from 2017 to 2020, but that trend stopped in 2021.

Then, were they changing at different rates? To answer this question, we calculated the increasing rates of real average wage between years using the formula:

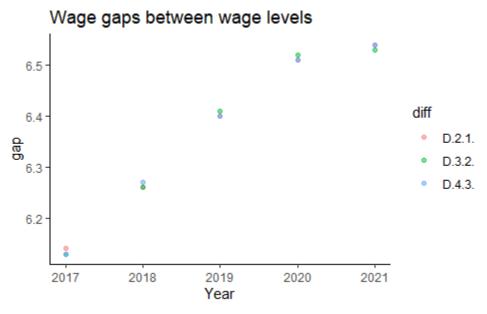
 $GrowthRate = [(RealAverageWage(Year2) - RealAverageWage(Year1)] \div RealAverageWage(Year1)]$  where Year2 - Year1 = 1.

Then, we visualized it again with a line plot.



It shows that they did change at different rates. For all wage levels, there were increasing growth rates in 2018 and 2019 then decrease in 2020 and 2021. Notice that except for level 4, all the 3 lower levels are expected to have negative growth rates in 2021, which means their real average wages are actually decreasing this year. Besides, if we label the absolute values for these 4 curves(Si, i = 1, 2, 3, 4), then the order remains almost the same: S1 > S2 > S3 > S4. This shows that the lower the wage level, the more drastic change in growth rate.

Let's see this from another angle. With all these differences, how did the gaps between wage levels change over time? Here, we used a scatter plot to illustrate the change of wage gaps between different wage levels across time.

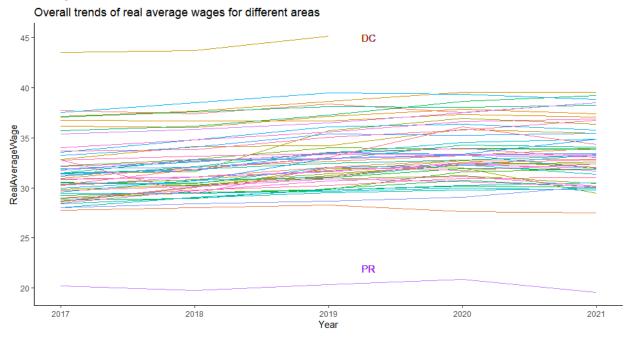


The overlaps of points show that wage gaps between two continuous wage levels are quite similar, given a certain year. But even with overlaps, it's clear enough for us to see that these gaps were becoming bigger

and bigger from 2017 to 2021. In other words, your gain from each promotion for the same occupation would increase during these years.

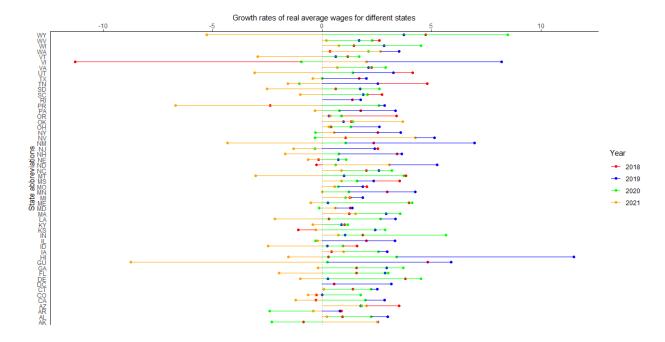
#### 3.2.2 Dimension 2: the geographical area(Area)

For different areas, the line plot of overall real average wages cannot really show us many details considering the total number of states.



But there are some interesting findings here. Firstly, the majority of states had constant increases in real average wages over time, though it's not the case for every state. Quite a few states experienced decreases in real average wages in 2020 and 2021. Secondly, except for the 2 outliers, the range of real average wages was between \$27.5 and \$40 per hour. Thirdly, about the 2 outliers, DC(District of Columbia) and PR(State of Puerto Rico). For historical and geographical reasons, it may not be that strange for PR to have this performance, but the curve of DC is quite surprising. We did expect DC to have higher or highest real average wages, but didn't know the gap would be this huge. Unfortunately, data for 2020 and 2021 is lost for DC, but the reasonable guess should still be the highest real average wages that outperform all other states enough to be an outlier.

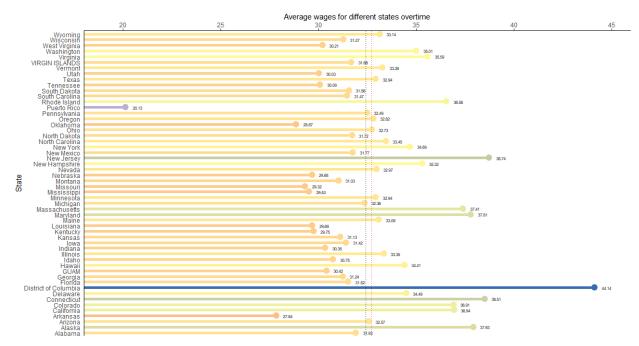
We dug deeper with growth rates.



We can see that for most of the states, there were positive growth rates from 2018 to 2020 and negative growth rates in 2021. Except for outliers, the range of growth rates should be within -5% and 5%. What's more, the average growth rate for all states seems to be highest in 2019. Focusing on outliers, GA(Georgia), NM(New Mexico), VI(Virgin Islands), WY(Wyoming) are states that had relatively unstable wages with big changes in growth rates.

Seeing from a bigger picture, we calculated the average prevailing wages over these years, using the formula below:

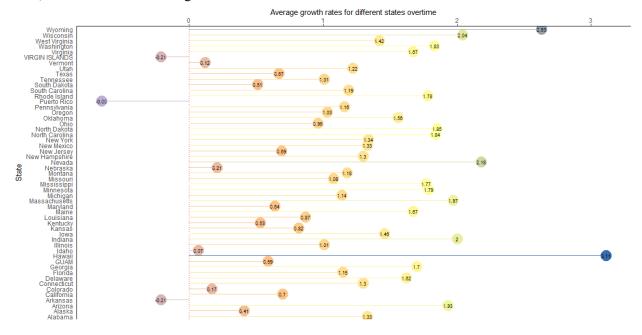
AverageWage = Average(RealAverageWage(Year i)) where i = 2017, 2018, 2019, 2020, 2021.



Here in this plot, the red dotted line is the mean of average wages for all states and the black one is the median.

As shown before, PR and DC are the outliers here. The majority of average wages stayed between \$30 and \$35 per hour, and the top 5 states with the highest average wages were District of Columbia, New Jersey, Connecticut, Alaska and Maryland.

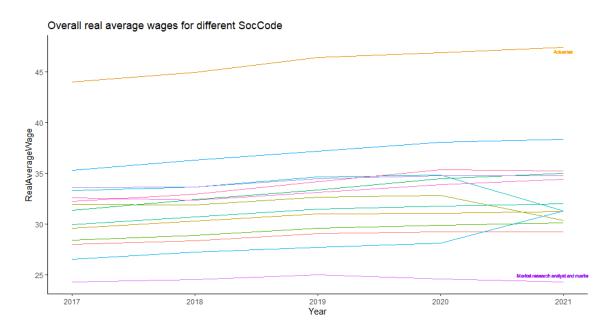
Then, let's look at the overall growth rates.



Despite those many negative growth rates in 2021, there were only 3 states that have negative overall growth rates during the past 5 years. For most parts of the country, average growth rates stayed between 0

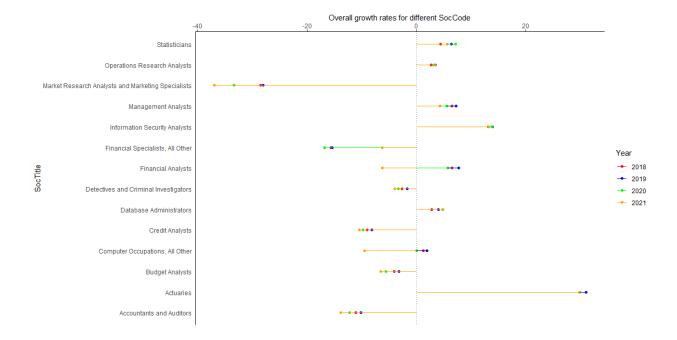
to 2%. The top 5 states with the highest overall growth rates in wages were Hawaii, Wyoming, Nevada, Wisconsin and Indiana.

# 3.2.3 Dimension 3: the title of the occupation(SocCode) Still, we started with the overall real average wages.



We can see that actuary seems to be the most promising occupation with the highest real average wages over years and constant increases, while market research analysts and market specialists tend to have the lowest income and experienced decreases in 2020 and 2021. Almost all occupations had increasing wages from 2017 to 2019, but the increasing rate slowed down or even turned negative in 2020 and 2021. Notably, financial analysts and computer occupations are expected to have a drastic decrease in real average wages in 2021, while we expect a big increase in wages for financial specialists in the same year.

Then, we calculated the growth rates each year.



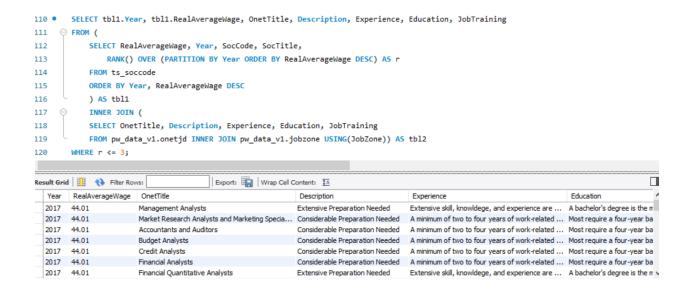
We included 14 different occupations defined by SocCode in our data set, and only 6 of them, which is less than half, had all positive growth rates from 2018 to 2021. The visualization of growth rate for different occupations are quite similar to the one of real average wages, with actuary and market research analysis and the marketing specialist as the highest and lowest outliers.

For detailed information about occupations, we need to come back to our database. We selected the top 3 occupations with the highest salaries each year. Then, linked it with detailed tables. One option of query is selecting top3 per year, then use 4 UNIONs to gather them together, since they have the same structures.

```
SELECT tbl3.Year, tbl3.RealAverageWage, OnetTitle, Description, Experience, Education, JobTraining
123 •
124

⇒ FROM (
125
       (SELECT RealAverageWage, Year, SocCode, SocTitle
       FROM ts_soccode
127
       WHERE Year = 2017
128
       ORDER BY RealAverageWage DESC
129
       LIMIT 3) UNION
130
     131
       FROM ts_soccode
       WHERE Year = 2018
132
       ORDER BY RealAverageWage DESC
133
134
       LIMIT 3) UNION
     ⊕ (SELECT RealAverageWage, Year, SocCode, SocTitle
135
140
     ⊕ (SELECT RealAverageWage, Year, SocCode, SocTitle
145
     FROM ts_soccode
147
       WHERE Year = 2021
       ORDER BY RealAverageWage DESC
148
       LIMIT 3)) AS tbl3 INNER JOIN (
149
150
          SELECT OnetTitle, Description, Experience, Education, JobTraining
151
          FROM pw_data_v1.onetjd INNER JOIN pw_data_v1.jobzone USING(JobZone)) AS tbl4
```

It's possible for 5 years, but this is not practical for more tables. So, we include another option of query here. We used PARTITION BY(), maped the RANK() over years and got the top3 using ORDER BY. The result tables are identical.



## 4. Summary

#### Combine business domain knowledge with technical skills

As we have mentioned in 3.1.3, jobs that require more technical skills tend to have higher wage rates. For business school graduates, we have the advantages in domain knowledge such as marketing, management or accounting, and it would be very beneficial to equip ourselves with more technical skills to gain an edge in the job market. It is recommended that we choose job titles that optimize our career interest, skillset and salary and benefits.

#### Geography is crucial in determining wage rate

Geographical area is an important factor in deciding the prevailing wage across the United States. Western coast and Eastern coast are obviously the highest-paid areas in the US, but there are also lots of other factors we need to consider when applying for jobs, such as living conditions, the cluster of industries, living costs and securities. Although people tend to have their preferences while choosing a place to settle down, we analyzed average prevailing wages and average growth rates of wages for different States to provide some basic information for them to make that decision. Interesting findings here are the outlier DC with really high prevailing wages, and states like Hawaii that may not seem to be very popular for job seekers but did have the highest growth rates in recent years.

#### Employees at higher wage level have more bargaining power

For a certain occupation, wages differ as employees' education level, professional experience and other soft skills vary. Our analysis shows that the growth rates tend to be more stable as wage levels become higher, and the wage gaps between levels are getting larger as well. This gives experts in higher wage levels more bargaining power. Job seekers can increase their wage rate from lower levels to higher ones through promotion or further education, however, it also takes time and money to reach higher levels. Therefore, it is worth thinking about the trade-off between loss and gain.

#### Is pandemic to blame for the decrease in wage rate?

If we see all the analyses horizontally, almost all decreases in wages happened in 2020 and 2021. The most reasonable guess for the explanation should be the COVID-19 pandemic. We could take further steps to research on the correlation between the pandemic and the decreases in prevailing wages.

Finally, refer to our plots and think about to what extent your occupation would be impacted by changes in the economic environment. Get prepared if you need to!