Capstone Project

**Analysis of wages in Colorado**

**James Flanagin**

**August 2019**

**Table of Contents**

**Introduction**

**Databases**

**Client**

**Data Wrangling**

**Data Exploration**

Wage and Metropolitan Statistical Area

Wage and Experience

Wage and Employee Count

Wage and Population

Job keywords and MSA

**Data Modeling (supervised Learning)**

Wages and Population, Experience, Employee

**Data Modeling (Unsupervised Learning)**

Logistic Regression

Job Segmentation

Principal Component Analysis

Word Frequency

**Assumptions and Limitations**

**Recommendations and Future Work**

**Conclusions**

**Introduction**

Colorado is one of the most desired places to live in the US. According to the US Census Beureu Colorado had the 5th highest total net migration in the US despite being the 22nd most populated state in the US. The nature and booming tech/ Cannabis industry have driven thousands of people to the state. The time frame we are going to explore is the growth from 2009-2015 which saw a positive migration of 484,379 new Colordan’s a 10% increase from 4,972,195 to 5,456,574 residents in 2015. The goal of this study is to see what the effect on wages looks like during an increase in population.’

**Databases**

<https://data.colorado.gov/Labor-Employment/Employment-Wages-in-Colorado/busm-qa5b>

Employment wages by occupation, year, and area, from Colorado Department of Labor and Employment (CDLE), 2009-2015

<https://data.colorado.gov/Labor-Employment/Population-Estimates-by-Year-for-Counties-in-Color/bu8h-8sux>

Annual population estimates for each year by county for the state of Colorado, from US Census Bureau, from 1900 to 2015 provided by the Colorado Department of Labor and Employment (CDLE).

The main Metropolitan Statistical Area (MSA) which are defined by the CDLE and featured in the provided dataset:

1. Pueblo MSA (Pink)
2. Greeley MSA (Brown)

3. Grand Junction MSA (Purple)

4 Fort Collins-Loveland MSA (Red)

5. Denver - Aurora MSA (Green)

6. Colorado Springs MSA (Orange)

7. Boulder-Longmont MSA (Blue)

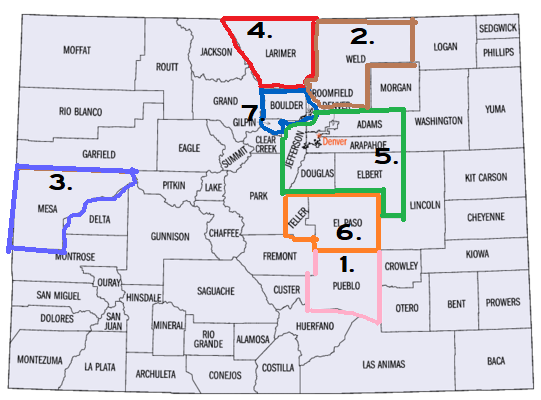


Fig1.

**Client**

There are several clients who would be interested in the outcome of this project

* A college Sophomore who is attending one of Colorado’s great universities who intends to stay in the state once they are finished. They could get a good feel for what occupation and location would be a great fit for them. Once they graduate they could have a basis for salary negation as well as how they should expect it to increase over the time period that they are working in a given job/ metropolitan area
* An Entrepreneur who is writing a business plan that is planning on setting up a business. Knowing what the mean salary given their projected business size as well as their office location will be key in understanding overhead. This can also be used for forecasting financial projections.
* Wages are the basis for standard of living in a given MSA. The city government can use this in instances of rent control and understanding tax revenue.
* Developers and residential real estate companies assessing rent and appraising property can learn what to expect from the average citizen of different MSA’s.

**Data Wrangling**

**Employment Wages:**

* A good deal of redundancies and useless columns were dropped i.e panelcode,
* any of the user defined columns were dropped as they presented conflicting information
* Any duplicate and summarized information. In order to avoid assessing a sum vs it’s parts I dropped any row that contained summarized information. i.e State Totals, hourly wage vs annual salary.
* After the Merge any descriptive data with one or few unique values ‘areatyname’, or any data that didn’t provide useful information eg. ‘indcodty’ was dropped

**Population Data:**

* Any summary data was dropped. Colorado and US totals were removed and County based data was all that remained
* Counties were grouped into their respective MSA
* Dates were adjusted to line up with Wage data (2009-2015)
* Columns that were useless were dropped. Population, periodyear and areaname were kept for merge purposes
* All county based data was summarized to reflect the total population of each MSA
* Only a portion of Colorado data is included in the major MSA’s. Counties not included were scrubbed.
* 2015 was missing data. Data was extrapolated based on the % change from prior year
* Data was merged based on Areaname

The final shape of the data= 19598 rows × 13 columns

**Data Exploration**

The final shape of the data= 19598 rows × 13 columns**.** There are 7 MSA’s as show above (Fig.1). Each row represents a unique year 2009-2015, and a unique occupation 836. This means that some occupations are missing a few years, or an MSA.



In the first part of analysis we look at the total Colorado change in average mean revenue.

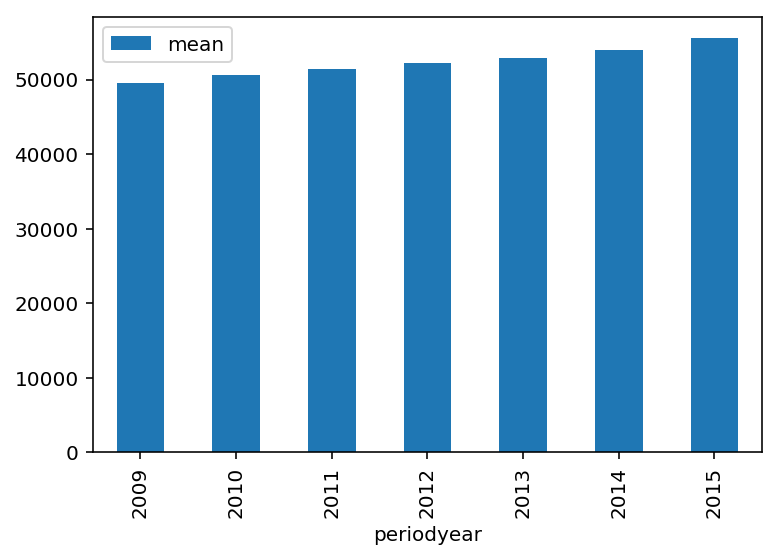


Fig.2 shows the man revenue by year is increasing slightly year over year on a steady trend

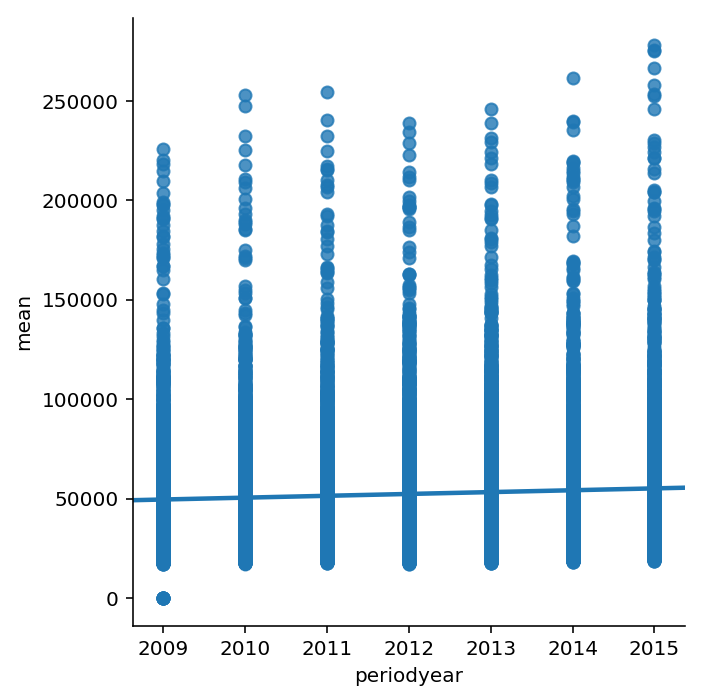


Fig.3 a lmplot shows that there are a few less outliers in 2012 and 2013 but still indicates a positive trendline for the state

When we divide it into MSA we can see a key distinction in mean revenue.

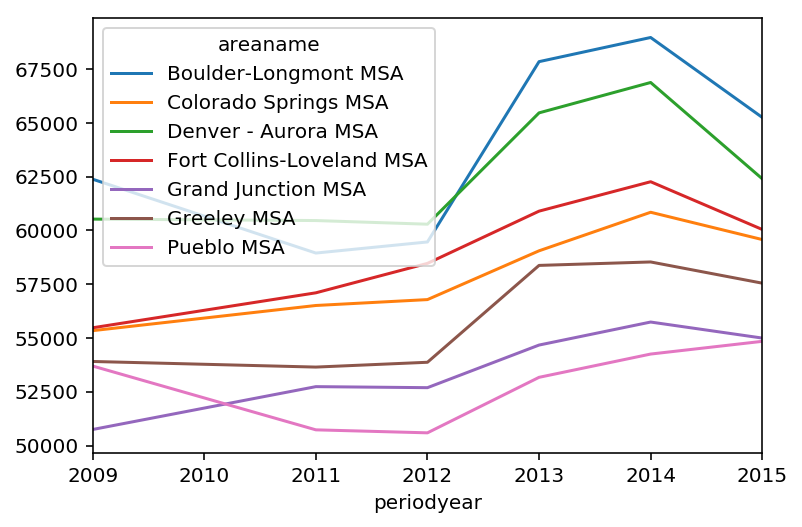


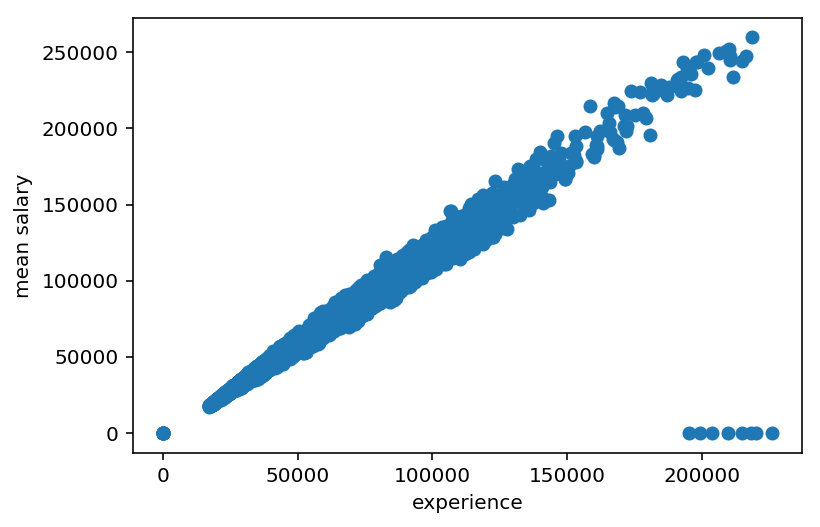
Fig.4 shows the trend in mean revenue across MSA. Boulder and Denver battle for first place as B-L pulls away in recent years. We show a slight dip in major locations except for Colorado springs, Fort Collins and Grand Junction (all agriculture heavy locations) after the recession in 2008 but increasing since 2012 in each location

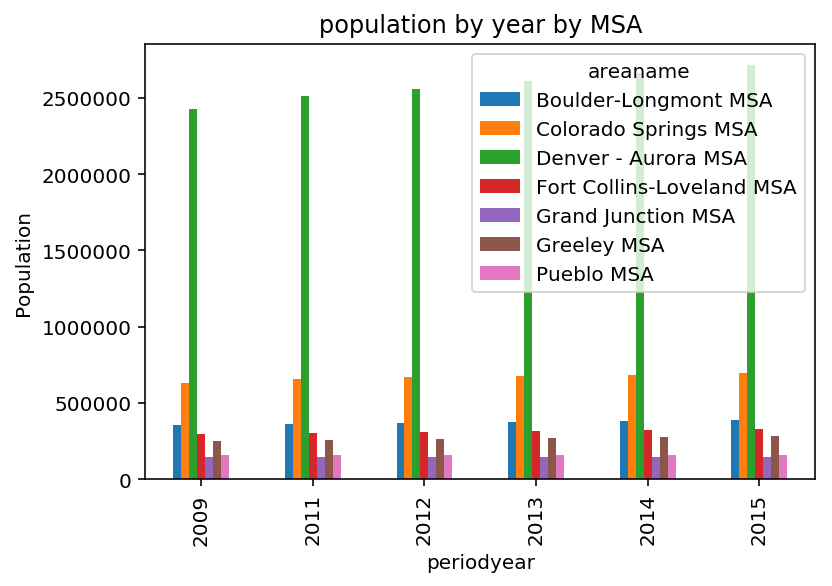
Fig.5 Compares experience and mean salary. We can see a tightly distributed positive correlation with some outliers where the mean salary is missing.

Fig.6 shows the population by year by MSA. This shows it steadily rising, with Denver-Aurora MSA being the clear leader. Colorado Springs is in second.

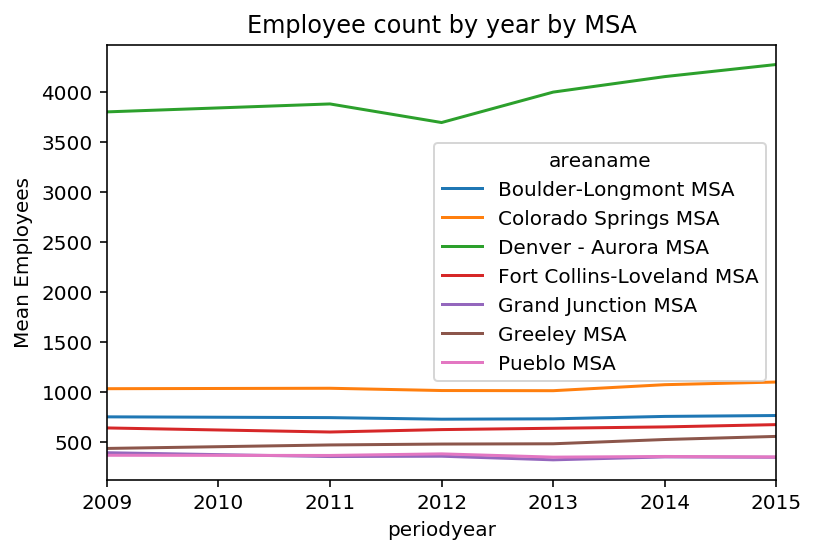


Fig.7 shows average employee count by locale. The more populated regions logically have higher concentration of employees by occupation.

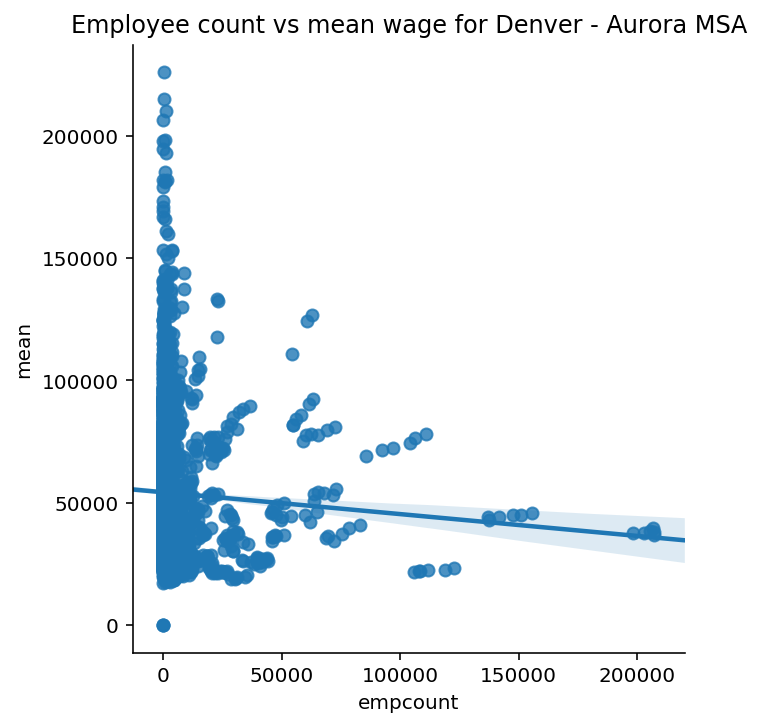


Fig.8 Shows Denver – Aurora MSA mean wage compared against the employee count for each occupation. Shows a slight negative correlation

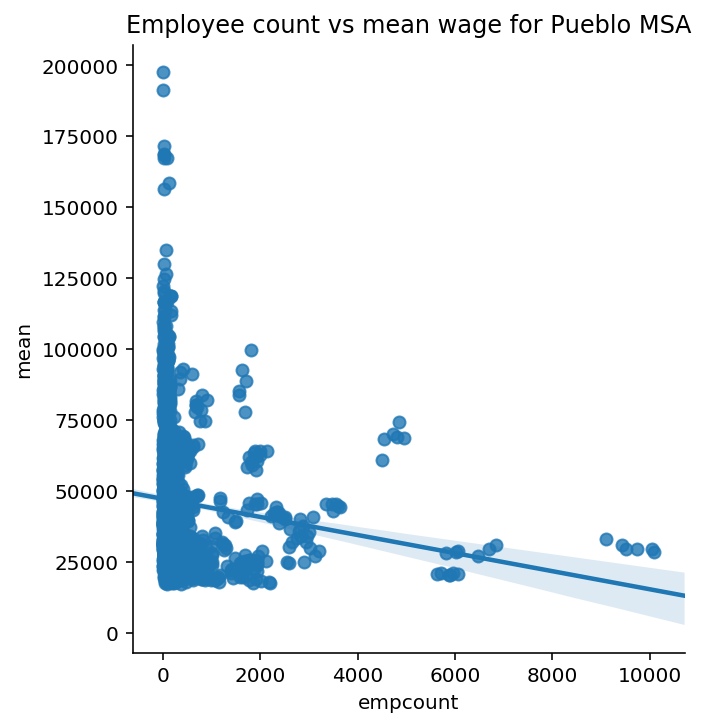


Fig.9 Shows Pueblo MSA mean revenue compared against the employee count for each occupation. Shows a higher negative correlation.

We can see that occupations with more workers tend to have a lower mean wage. We also see that across a high population area(Denver), and a low population area(pueblo) that the majority of jobs have less than 2000 total workers.

Finally let’s look at the ceilings and floors for each job by state

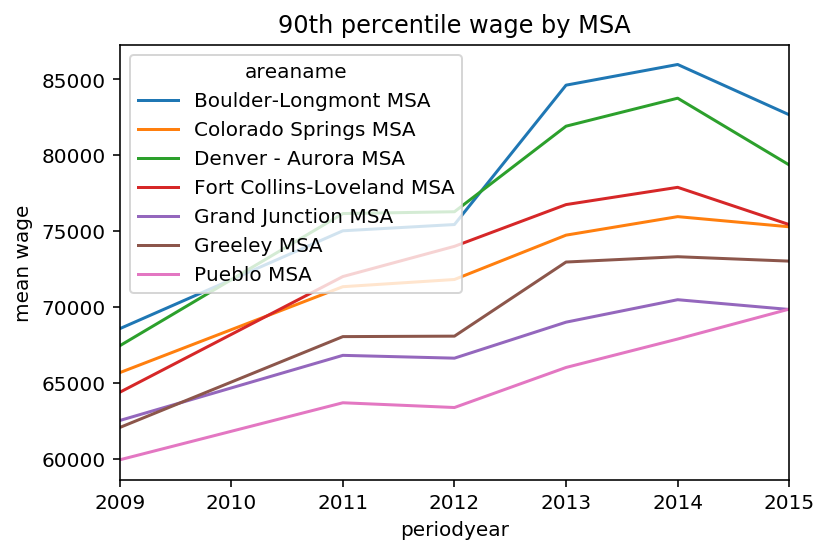


Fig.10 show’s the 90th percentile or the mean top pay by year for each city. Boulder-Longmont leads the pack and is $25k higher in the peak year of 2014. We can see they all follow similar trends, and were relatively stable in a recovering economy

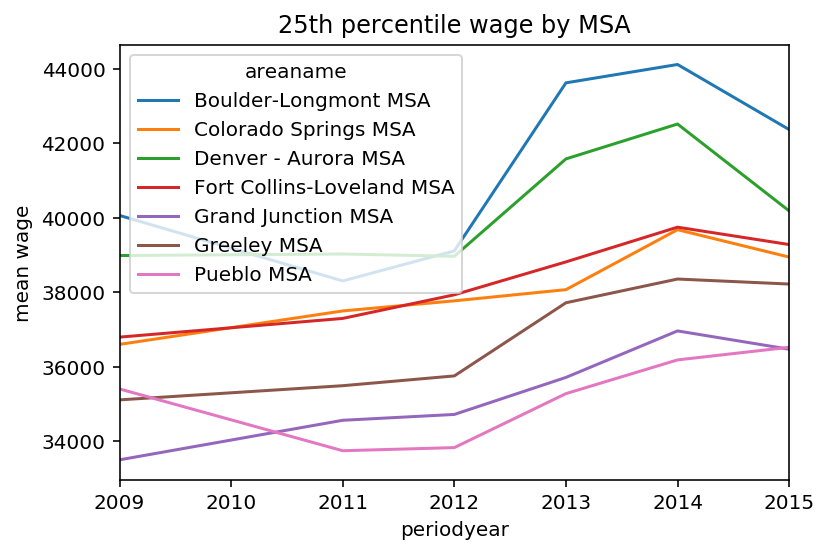


Fig.11 shows the bottom 25th percentile of wage by MSA. We see that the lower paying parts of each occupation are much slower to recover from the recession. They have a consistent negative trend until 2012 where each MSA increases across