The Effect of Experience Level on Salary

Ido Terner and **Eyal Finkelshtein**
Industrial Engineering and Management Department
The Technion - Israel Institute of Technology
Haifa 32000, Israel
{ido.terner, eyal.f}@campus.technion.ac.il

Abstract

Many people all over the world invest many years studying in universities and colleges, and gaining experience in a profession under the assumption that it will **cause** in better position, conditions, and salary. Indeed, we see a correlation between the two, but as the fundamental saying in the field of causality goes - **correlation doesn't imply causation**. In this project, we will check whether this claim is indeed true. We perform an inference on the **"Data Science Job Salaries"** dataset. For estimation of the causal effect, we estimate the propensity scores, perform Nearest Neighbor Propensity Score Matching (PSM) algorithm, and compute the Average Treatment Effect (ATE), Average Treatment effect on Treated (ATT), and Average Treatment effect on Controlled (ATC).

1 Introduction

Making predictions about the causal effects of actions is a major problem in many domains. For example a doctor deciding which medication will cause better outcomes for his patients, or a teacher deciding which task would most benefit the class. In this project, we focus on "High-Tech" environment, and in particular data science jobs. There are several "High-Tech" experience levels - entry, intermediate, senior, and expert - each one represents the seniority and the achievements in the profession. Entry level is the lowest experience level and expert is the highest. The most natural thing to ask about this setting is whether there is an effect of the employee's experience level on his salary. One might think that it is obvious that with more experience comes more money, but as known, correlation doesn't imply causation. This question is very important, since people all around the world are spending a lot of time and energy finding jobs and keeping them to gain experience, even if they don't like or enjoy practicing their profession. If this claim is false, we can assume that getting a higher salary is only a matter of luck.

First, in section 2 we present and give a brief explanation of the used methods. In Section 3 we describe the dataset we use in this project and discuss the assumptions we make. Section 4 presents the experiments we have conducted, and Section 5 presents the results, which are followed by possible weaknesses and discussion in Sections 6, 7 respectively.

The code used to generate the results presented in this paper can be found on GitHub[1].

[1] https://github.com/EyalFinkel/experience_effect_on_salary

2 Methods

In this section, we present and explain the methods used in the project.

2.1 Propensity Score Estimation

The propensity score represent the probability of treatment assignment to a unit, given observed baseline characteristics (observed covariates). The propensity score allows to design and analyze an observational (non-randomized) study so that it mimics some of the particular characteristics of a randomized controlled trial. The probability of getting treatment conditioned on the covariates is calculated using the formula $P(X_i) = P(T_i = 1|X_i)$ where T_i, X_i are the treatment and the covariates of unit i respectively. The propensity computation is based on running a logistic regression model on the treatment indicator on functions of the covariates. For this project, we compute the propensity using all the linear covariates. Other methods may use different functions of the covariates, such as quadratic functions ($X^2(j)$, $X(j)X(k)$, etc.).

2.2 Trimming Common Support

The common support is the intersection of the support of the propensity score distribution for the treated and the control groups. After the propensity score estimation, we need to trim the set of observations to the common support of the propensity score distribution [1]. It is done simply by removing the observations from each group that are outside the common support. This step is preformed because for observations outside the common support, we can't find a matching observation with a similar propensity score, which is crucial for the propensity score matching.

2.3 Propensity Score Matching

When estimating causal effects using observational data, it is desirable to replicate a randomized experiment as closely as possible by obtaining treated and control groups with similar covariate distributions. Propensity Score Matching (PSM) [2] comes in handy for this purpose.

PSM is a method used to approximate experimental results to recover the causal effect from observational data. This method attempts to reduce the bias due to confounding variables that could be found in an estimate of the treatment effect obtained from simply comparing outcomes among units that received the treatment versus those that did not.

The 1 nearest neighbor matching algorithm [3] is presented in Algorithm 1.

Algorithm 1 1 Nearest Neighbor Matching

Require: data X, distance metric $d(\cdot,\cdot)$, and treatment indicator T
  for i **do**
    $J(i) = \text{argmin}_{j:t_j \neq t_i} d(x_i, x_j)$ ▷ $J(i)$ is the nearest counterfactual neighbor of i
    if $t_i = 1$ **then** ▷ If the case is treated
      $\widehat{ITE(i)} = y_i - y_{J(i)}$
    else ▷ If the case is not treated
      $\widehat{ITE(i)} = y_{J(i)} - y_i$
    end if
  end for
  $\widehat{ATE} = \frac{1}{n} \sum_{i=1}^{n} \widehat{ITE(i)}$ ▷ n is the number of samples

3 Dataset

3.1 Raw Data

The dataset we used for this project is "Data Science Job Salaries"[2]. It contains 609 rows, where each row represents an employee. The properties and their possible values, that are used for this project, are:

- **Work Year**: The year the salary was paid - 2020-2022.
- **Experience Level**: The experience level in the job during the year - EN entry, MI intermediate/mid, SE senior, EX expert/director.
- **Employment Type**: The type of employment - PT part-time, FT full-time, CT contract, FL freelance.
- **Job Title**: The role worked in during the year - Data Scientist, Data Engineer, Data Analyst, Machine Learning Engineer and Research Scientist.
- **Salary in USD**: The salary in USD.
- **Employee Residence**: Employee's primary country of residence during the work year as an ISO 3166 code - US United States, GB Great Britain, etc.
- **Remote Ratio**: The overall amount of work done remotely - 0 No remote work, 50 partially remote, 100 fully remote.
- **Company Location**: The country of the employer's main office as an ISO 3166 code - US United States, GB Great Britain, etc.
- **Company Size**: The average number of people that worked for the company during the year - S less than 50, M 50-250, L more than 250.
- **Salary Currency**: The currency of the salary paid as an ISO 4217 currency code - USD US dollar, EUR Euro, etc.
- **Salary**: The total gross salary amount paid in the local salary currency.

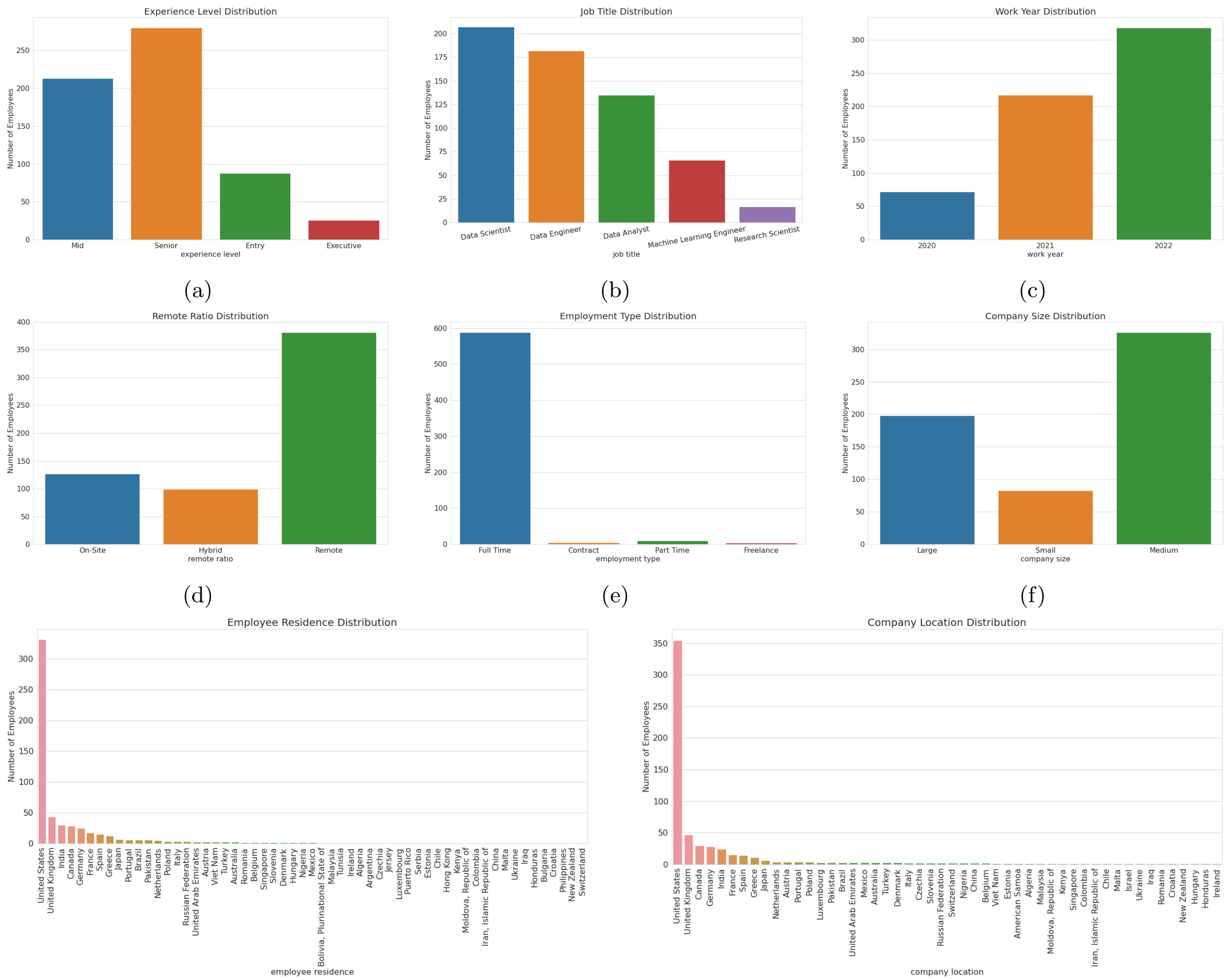
3.2 Pre-process

As a first step of pre-processing, the features for each entry were chosen. The choice of features made with the confounders in mind, to be able to achieve meaningful results. The columns we kept are: Work Year, Experience Level, Employment Type, Job Title, Salary in USD, Employee Residence, Remote Ratio, Company Location, and Company Size. The Salary Currency and Salary columns were dropped, since they don't offer a uniform measure to compare salaries of different employees, and the Company Location and Employee Residence were already used as the geographic confounders.

Later, different job titles that represent the same position, such as ML Engineer and Machine Learning Engineer, were processed to avoid such duplications. We grouped the job titles into five different groups, where each one represents a collective of jobs - Data Scientist, Data Engineer, Data Analyst, Machine Learning Engineer and Research Scientist. In addition, since most of the raw data contains strings, we converted each unique string to a different numerical value.

[2] https://www.kaggle.com/datasets/ruchi798/data-science-job-salaries

Figure 1 shows the distributions of all the categorical features. The distribution of of the salary in USD, which is the outcome, is presented in Figure 2.



(a) (b) (c)

(d) (e) (f)

(g) (h)

Figure 1: Distributions of all categorical features: experience level (a), work year (b), job title (c), remote ratio (d), employment type (e), company size (f), employee residence (g), company location (h).

3.3 Assumptions

To estimate the causal effect using our data, we have to make several assumptions on the data.

The **Stable Unit Treatment Value Assumption (SUTVA)** holds since each row in the dataset represents an independent person, and the individuals in the dataset are scattered over many countries and companies. Moreover, the experience level (for example junior and senior levels) are pretty standard across companies and countries, so there is no difference between the experience level definitions in different companies. These also allow us to assume **consistency**, since we know that each individual

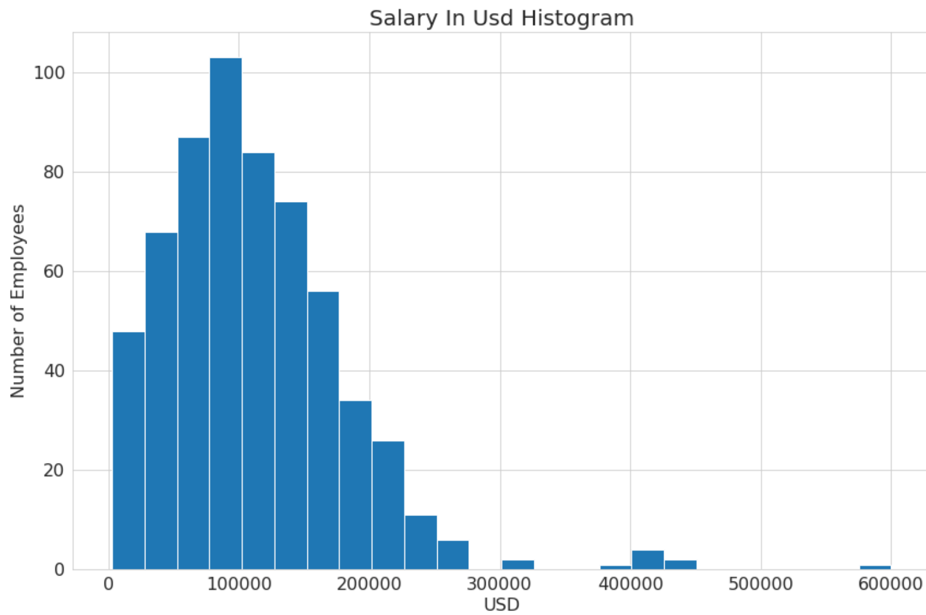


Figure 2: Histogram of the salaries in the dataset. All salaries are converted to USD.

will get higher salary with accumulation of experience.

The **common support (overlap)** assumption, which is $\forall t, x : P(T = t|X = x) > 0$, initially doesn't hold for the dataset, but is fixed by trimming the data, which makes the assumption hold.

One assumption that doesn't necessarily hold is the **ignorability** assumption, which says that all the existing confounders are measured. In this project, we assume that other than the confounders we named earlier in this section, there are no more confounders. We think that it is a fair assumption since we named many relevant confounders. In theory, there can be many hidden confounders, which have an indirect affect on the treatment and on the outcome, but after a long consideration we came to the conclusion that ignorability can be assumed for this dataset.

4 Experiments

We have conducted several experiments, that differ in the division of the experience levels into treated and control groups. All experiments follow the same pattern:

1. Divide the data into two groups of treated and control, by a certain criteria.
2. Estimate the propensity score, and apply the matching algorithm (see Section 2).
3. Compare the distributions of the propensity score and the covariates for the treated and the control groups before and after the matching.
4. Calculate the causal estimators ATT, ATC, and ATE, each one with a 95% confidence interval.

As mentioned in Section 3 and as can be seen in Figure 1a, the experience level is divided into 4 groups - Entry (EN), Mid (MI), Senior (SE), and Expert (EX). For each experiment we include different experience levels in the control group (which is considered to have low experience level) and in the treated group (which is considered to have high experience level). The distributions of treatment for all the experiments are shown in Figure 3.

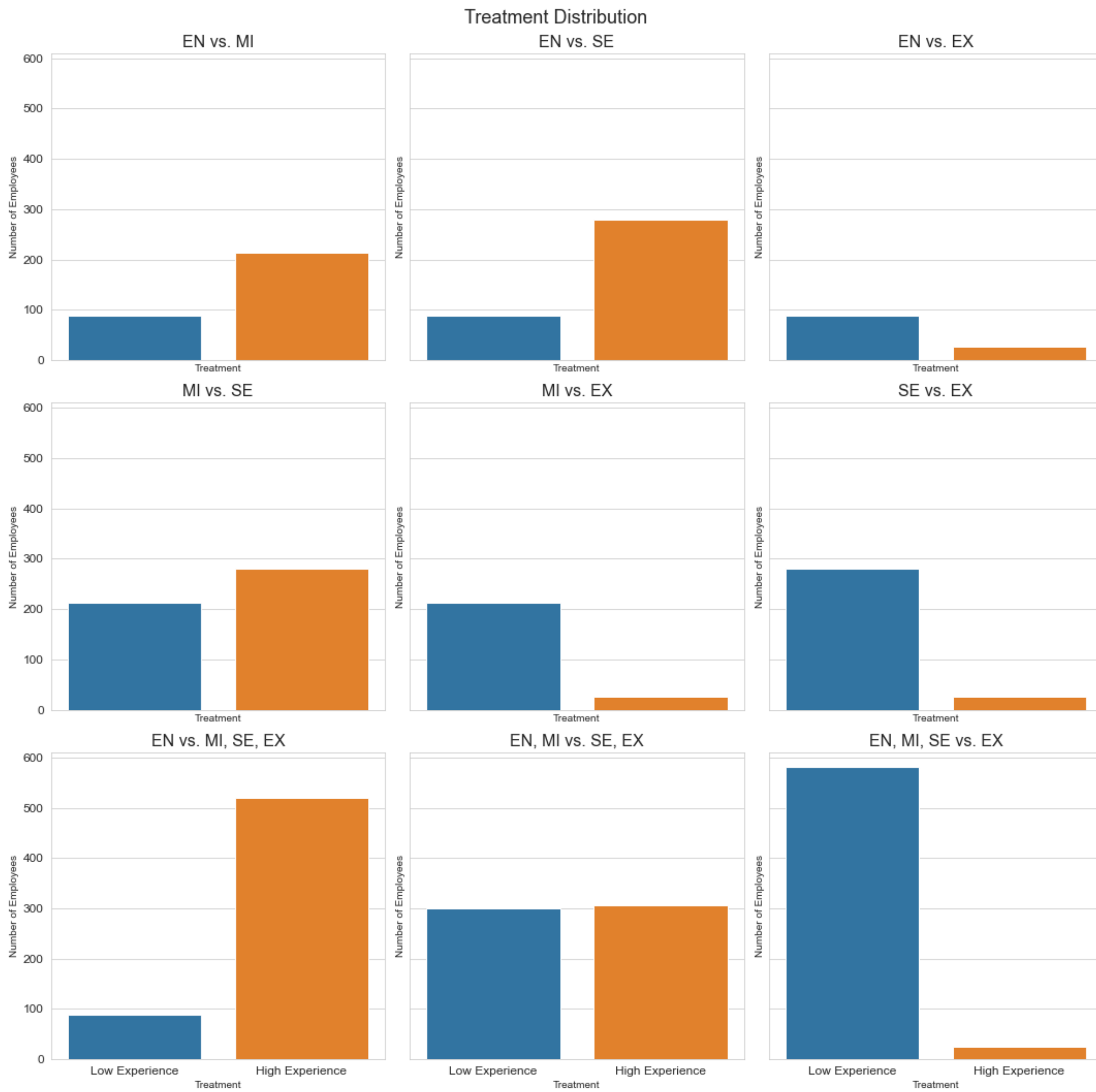


Figure 3: Distributions of treatment for all experiments. From left to right, and top to bottom: Entry vs. Mid; Entry vs. Senior; Entry vs. Expert; Mid vs. Senior; Mid vs. Expert; Senior vs. Expert; Entry vs. Mid, Senior, and Expert; Entry and Mid vs. Senior and Expert; Entry, Mid, and Senior vs. Expert.

5 Results

In this section, we provide a detailed experiment analysis of our model on the Data Science Job Salaries dataset. we will begin by shortly describing the causal estimators we used in this project, and then describing the results for each experiment.

5.1 Causal Estimators

Since our project tries to solve the question whether an employee's experience level affects their salary, we want to compute causal estimators that will give us an indication of the treatment affects. To do that, we have used three causal estimators - Average Treatment Effect (ATE), Average Treatment effect on Treated (ATT), and Average Treatment effect on Controlled (ATC).
ATE is the expected causal effect (salary difference) of the treatment (experience level) across all individuals in the population, and is estimated by the difference in average outcomes between units assigned to the treatment group (high experience) and units assigned to the control group (low experience):

$$ATE = \mathbb{E}[Y_1 - Y_0]$$

ATT is the expected causal effect of the treatment for individuals in the treatment group, and is given by:

$$ATT = \mathbb{E}[Y_1 - Y_0|T = 1]$$

ATC is the expected causal effect of the treatment for individuals in the control group, and is given by:

$$ATC = \mathbb{E}[Y_1 - Y_0|T = 0]$$

Notice that $\mathbb{E}[Y_0|T = 1]$ and $\mathbb{E}[Y_1|T = 0]$ are not observed in the experiments. As mentioned before, for each causal estimator, we calculate a confidence interval of the effect with a confidence level of 95%.

5.2 Experiment Results

Figures 4 and 5 show the propensity score, job title, company size, and work year before and after the matching, for the Entry and Mid vs. Senior and Expert and the Entry vs. Mid experiments respectively. We can see that the Propensity Score Matching indeed gets us closer to replicating a randomized experiment, to reduce the bias.

Experiment	$\widehat{\textbf{ATT}}$ [K$]	$\widehat{\textbf{ATC}}$ [K$]	$\widehat{\textbf{ATE}}$ [K$]
Entry vs. Mid, Senior, Expert	52.017 ± 6.859	42.706 ± 18.089	50.637 ± 6.433
Entry, Mid vs. Senior, Expert	21.972 ± 9.053	50.019 ± 8.889	35.879 ± 6.443
Entry, Mid, Senior vs. Expert	84.257 ± 52.376	106.011 ± 11.019	104.645 ± 10.850
Entry vs. Mid	24.201 ± 10.435	41.397 ± 21.954	29.279 ± 9.844
Entry vs. Senior	54.863 ± 8.621	51.076 ± 15.805	53.973 ± 7.571
Entry vs. Expert	100.718 ± 46.539	108.140 ± 21.783	106.171 ± 20.223
Mid vs. Senior	20.612 ± 8.542	32.866 ± 10.534	25.835 ± 6.669
Mid vs. Expert	94.819 ± 44.135	64.909 ± 13.371	68.444 ± 12.956
Senior vs. Expert	65.326 ± 43.166	24.427 ± 10.937	28.486 ± 10.844

Table 1: The numerical results of all causal estimators - ATT, ATC, and ATE, for each experiment, with 95% confidence interval.

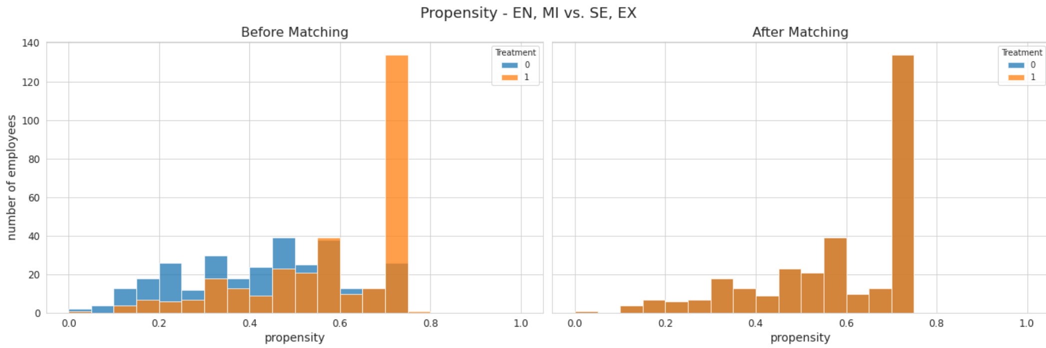
The ATC, ATT, and ATE were calculated for all experiments to answer the question at hand. Table 1 shows the numerical results for there estimators, together with a 95% confidence interval. Figure 6 shows the same results in plots, to ease the comparison between the experiments and the final deduction. It is clear from Figure 6 that with significance level 0.05, **there is an affect of the experience on the salary**, since all confidence intervals don't include 0. Notice the large ATT confidence intervals for the experiments that use only the Expert level as a treatment group. Since these experiments consider only the treatment group, and the Expert group contains relatively small number of employees (as can be seen in Figure 1a), we get a large confidence interval, which is inversely dependent on $\sqrt{n}$. The same phenomenon is observed in the ATC confidence intervals for the experiments that use only the Entry level as a control group, but less evidently, since the Entry group contains more employees. From the ATE results we can infer that the salary difference between every two adjacent experience levels is around 30,000$.

6 Possible Weaknesses

While working on the project, we tried as much as possible to avoid making unnecessary assumptions and ignore things, such as more unmeasured confounders, for our results to be correct. Unfortunately, there are cases where we had to do so, and therefore they can be a possible weakness. One of them is the number of samples we have. Since our dataset is relatively small (only 609 entries), it might have affected our results, since we may not have enough data for each group (treated and control) in each experiment. In addition, since we assume there are no more confounders except the ones given in the dataset, we could get inaccurate results, in case there are in fact confounders we couldn't measure. In this case, the ignorability assumption won't hold, and the results we got won't be accurate.

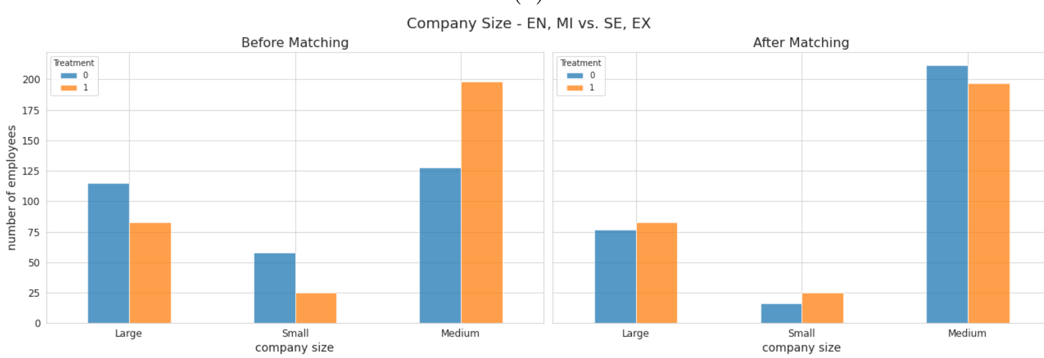
7 Discussion

This project tries to answer the causal question of whether there is an effect of the employee's experience level on their salary. After presenting and looking at the results, we conclude that the null-hypothesis can be rejected, and that experience indeed affects the salary. Every experiment we conducted showed a significant salary gap, with a confidence level of 95%. Judging by these results, we can say that investing time and energy in gaining experience is an important step towards increasing the paycheck in the data science professions. In addition, we would like to conduct this experiment for jobs from other domains and professions in the future, to better understand whether experience affects salaries.



(a)

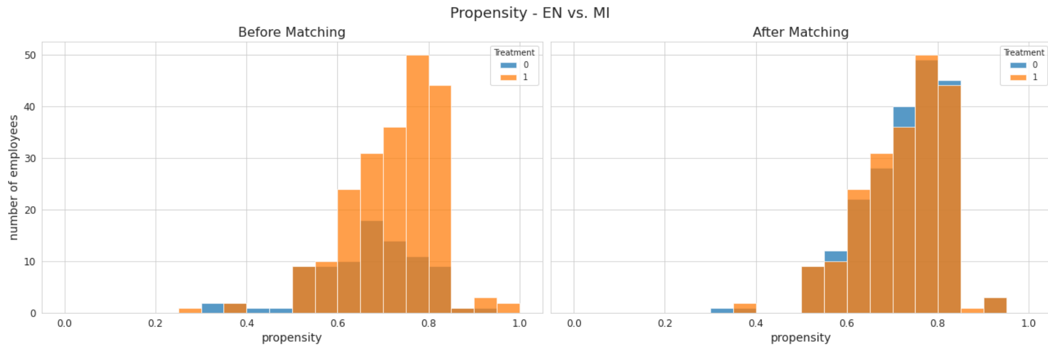
(b)



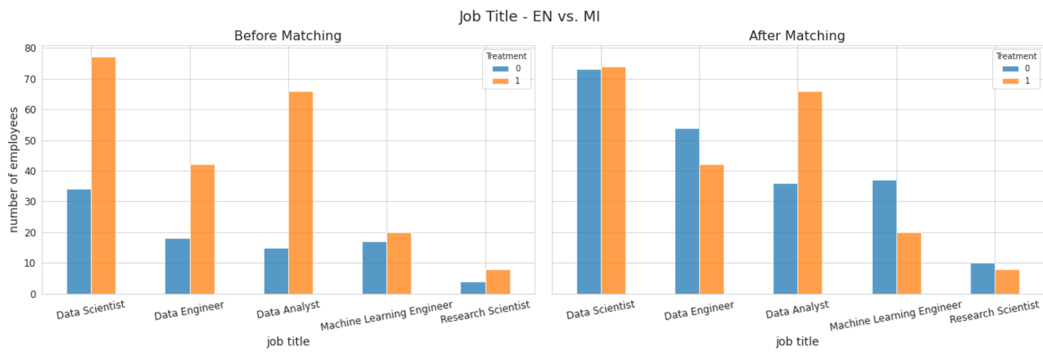
(c)

(d)

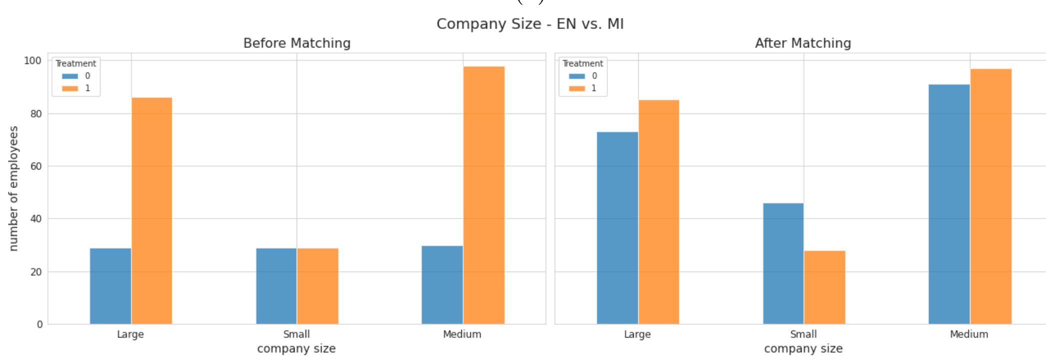
Figure 4: The distribution of propensity score (a) and some of the features (b)-(d) before and after matching, for the Entry and Mid vs. Senior and Expert experiment.



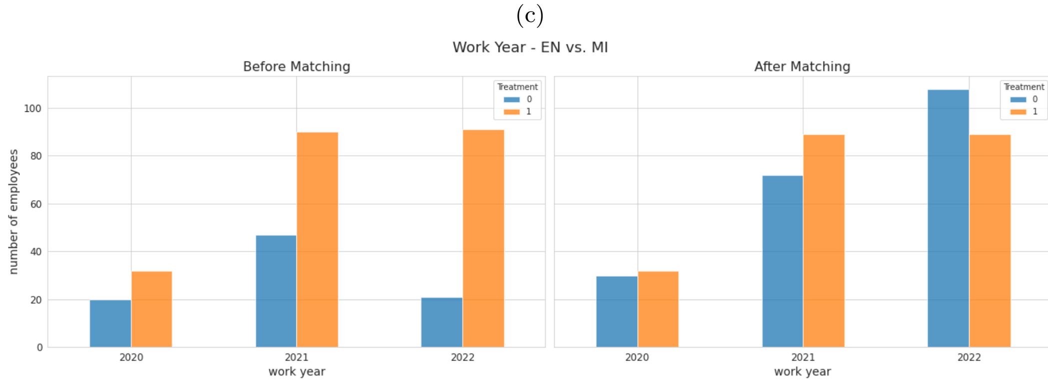
(a)



(b)

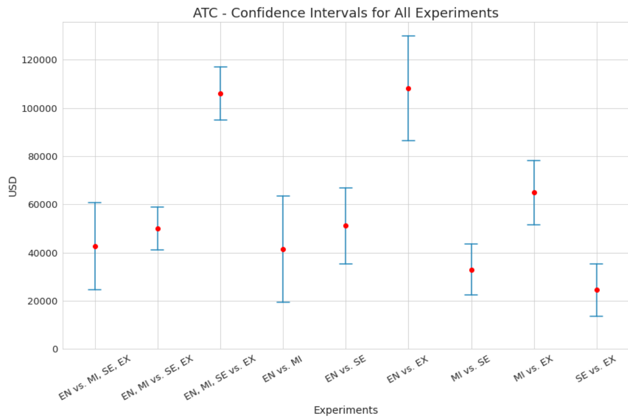


(c)

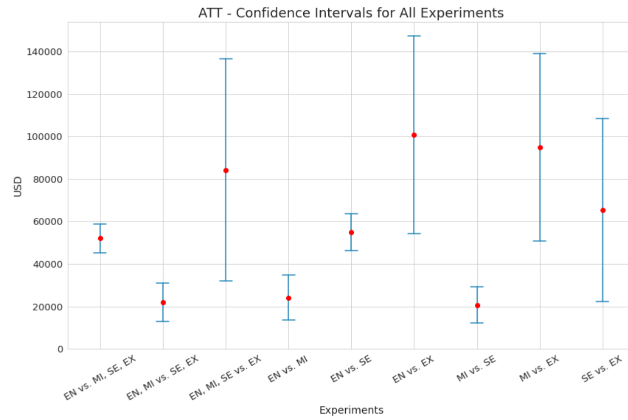


(d)

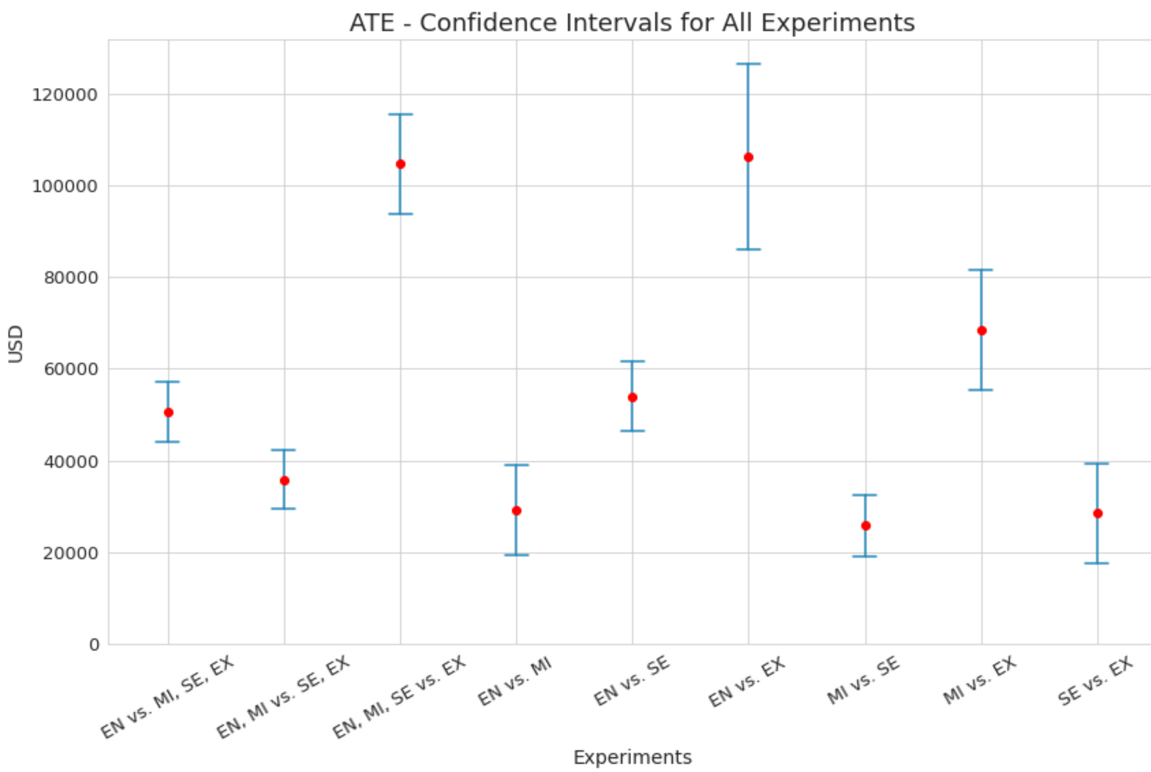
Figure 5: The distribution of propensity score (a) and some of the features (b)-(d) before and after matching, for the Entry vs. Mid experiment.



(a)



(b)



(c)

Figure 6: Confidence intervals with confidence level of 95%, for all experiments, and for each of the causal estimators - ATC (a), ATT (b), and ATE (c).

References

[1] Til Stürmer, Michael Webster-Clark, Jennifer L Lund, Richard Wyss, Alan R Ellis, Mark Lunt, Kenneth J Rothman, and Robert J Glynn. Propensity Score Weighting and Trimming Strategies for Reducing Variance and Bias of Treatment Effect Estimates: A Simulation Study. *American Journal of Epidemiology*, 190(8):1659–1670, 02 2021.

[2] Guido W. Imbens and Donald B. Rubin. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press, 2015.

[3] Bradley Efron and Robert J. Tibshirani. *An Introduction to the Bootstrap*. Number 57 in Monographs on Statistics and Applied Probability. Chapman & Hall/CRC, Boca Raton, Florida, USA, 1993.