2019

Optimize the Effectiveness of Recruiting Campaigns

Ryan A. Talk
Southern Methodist University, rtalk@smu.edu

Lakshmi Bobbillapati
Southern Methodist University, nbobbillapati@smu.edu

Marshall Coyle
Capital Group, msrc@capgroup.com

Optimize the Effectiveness
of Recruiting Campaigns

Nagalakshmi Bobbillapati [1], Ryan Talk [1] and Marshall Coyle [2]

[1] Southern Methodist University, Dallas, TX 75275 USA
[2] Capital Group, 6455 Irvine Ctr Dr, Irvine, CA 92618 USA
`{nbobbillapati, rtalk}@smu.edu, msrc@capgroup.com`

Abstract. Recruiting marketing plays an important role in the talent acquisition strategy today. To find the best candidates, companies make substantial investments through numerous recruiting agencies, job boards, and internal systems such as Indeed, LinkedIn, Monster, and Talent Communities. In this paper we obtained a company's LinkedIn Job Posting data to try to predict the number of visits they will receive for each job posting based on the time of the year it is posted. Developing a model to predict the number of visits or applicants has the potential to optimize a recruiter's budget spending while still meeting their recruiting needs. This in turn can save the company money. We pursued a classical time series approach to lay the foundation for future research on this dataset as well as to create a quick deliverable model for the company.

Keywords: Time Series, Business Intelligence, Data Science, Human Resources, Talent Acquisition.

1 Introduction

Great talent is a powerful and consistent source of competitive advantage in business [1]. To obtain this competitive advantage, companies are investing in recruitment marketing. Recruitment marketing consists of the methods that a company employs to reach out and attract qualified candidates and get them to apply for position at their company. These campaigns can result in both active (who are actively looking for job) and passive (who are not actively searching for job) candidates.

There are several common approaches to Recruiting marketing: Social Media, Email Campaigns, Jobs Ads, and improving the companies Career site. The primary focus being Social Media. The goal for these several approaches is to attract, persuade, connect, and engage with the clients with the desired result of hiring the top talent.

A study conducted by the Society for Human Resource Management (SHRM) claims that it costs on average $4,129 and takes on average 42 days to fill a new position. Another source, Glassdoor, claim a $4,000 cost to hire a new employee and an average of 52 days to fill a position. When a company is trying to find a replacement for an existing position, the costs become even greater to between 50 – 60% of the previous employee's annual salary [1].

2

Companies typically spend millions of dollars on branding and recruiting marketing to hire the right candidates [2]. With the high dollar amount spent for talent acquisition, a concern for companies can be their Return on Investment (ROI). Are jobs posted on the right job boards to get the right candidates? How much money is invested on job boards? Which job boards are getting the qualified candidates? This paper focuses on optimizing the effectiveness of the common approaches (listed above) in attracting, engaging and hiring the quality candidates; comparing the cost per hire vs the cost of the employee; and providing a metric to measure the ROI from these campaigns. As part of this project, we are going to explore data for one such company and predict the number of visits / subscribes that they can get based on the job they have posted.

2 Analytics and Recruiting Marketing

The way that businesses and recruiters operate is drastically changing due to data. Many companies are pushing their marketing operations towards data-based decision making to become more successful. In order to optimize a client's return on investment, recruiters are being pushed to demonstrate their value. Big data is a natural way to showcase value, thus increasing the demand for data specialists in this area [3].

In order to extract the most value or ROI from data, a large company requires a combination of both data science and marketing analytics specialists. According to Market Pro Inc, the key nuance that distinguishes Market Analytics from data science is the tactical nature of market analytics vs the strategic nature of data science. Marketing analytics consultants tend to be more reactive to what has happened and why. This is important to understand with regards to the ROI for marketing initiatives [4].

Data scientists on the other hand are more proactive in predicting the future based on upcoming seasonal cycles, potential competition from competitors, demand spikes etc. They can also anticipate and optimize the effectiveness of the campaigns, expected number of candidates, and potential hires. They can also build some machine learning techniques to help with future prediction of which job board can yield the best candidates and when. Areas of recruiting where data science can make a difference is discussed in the following sub sections.

2.1 Ability to attract the right talent

A Recruiter's job would change if they know ahead of time the time it would take to fill a position based on market trends. What are the similar jobs posted in the geographical location? Which job boards can yield the most candidates and what would cost per hire look like? Based on these recruiters can set the right expectations with the hiring team on the likelihood of finding right talent given the job description and salary.

2.2 Operational Strategy

Based on the data, recruiters can now hold a seat at the table as the companies define their operational strategy. They can recommend the likelihood of hiring the resources needed based on the geographical area and the skills that can be obtained around the area. Also, determine how much budget needs to be set aside to find the right talent.

2.3 Planning & Forecasting

Based on the hiring from previous years and business roadmaps, recruiters can forecast the headcount and skillset needed to fulfill the business needs. Conversely, a talent-rich location will give the business the confidence that it can deliver on the customer's expectations.

Recruiters and companies utilize job boards to advertise open positions and scan resume databases [5]. Job seekers use job boards to find new career opportunities and apply for jobs online. There are free job boards and premium job boards. Premium Job boards are commonly referred to as a source to find "active job seekers "and have advanced networks or offer other features which is why most companies are invest in them. Whereas, free job boards where the jobs are displayed based on the date posted. Our paper will focus on validating if time plays any factors in getting the number of visits and applies from candidates through these job boards. In our dataset we see several such job boards such as LinkedIn, Glassdoor, Monster, Ziprecruiter and others. During the preliminary analysis it showed that most of the candidates' visits came from LinkedIn. Also based on the domain knowledge, we focused our research on LinkedIn at this time. Based on the time series model, we are trying to predict the number of visits expected for a job posting based on the job role and functional area. Also, predict the best month to post the job on LinkedIn to get the most visits / applies. We can even enhance our model to granular level by looking at the education level and experience level needed for the job.

3 Recruiting Campaign Dataset

The raw dataset consists of data from the Applicant Tracking System (ATS), the Recruiting Marketing (RMK) system, and internal job posting data to create a comprehensive dataset that describes the voyage from a candidate's first visit to their eventual hire [6].

3.1 Importance of the Data

This combined dataset assists in answering questions and driving metrics that matter to the Talent Acquisition team related to the following:

- Marketing: How do we structure future communications/campaigns?
- Candidate Experience: What is our applicant drop-off rate? How is our Talent Community mix?

4

- Continuous Improvement: Where should our efforts be focused on with our sourcing strategy?
- Data & Analytics: Uncover insights to increase efficiencies and strengthen strategy

Recruiting marketing gathers, distributes and markets our jobs to candidates. RMK accomplishes this by investing in job boards such as LinkedIn, Glassdoor, Indeed and others. The data collected from these investments gives us information about our candidates, job applications, and when/how/what sources the candidates came from.

Through this RMK dataset we have access to the following:

- Visitors: How many unique visitors came to your site
- Members: Individuals who become members of the Talent Community by signing up as a member or applying for a job
- Applies: Individuals who have clicked "Apply" in Recruiting Marketing and were handed off to the applicant tracking system.

Now, the Applicant Tracking System manages the recruiting and hiring process which include the job posting and job applications. Its primary purpose is to allow anyone involved in the hiring process to stay connected on the progress of filling a job. These people include candidates, hiring managers, HR representatives, and recruiters. The data gathered in this system involves the number of completed candidate applications, number of interviews, number of qualified candidates, number of offers, and number of hires. It is also a living dataset that is updated on a weekly basis. Lastly, the job posting data contains information about the specific job posted. This includes job description, job title, education level, experience level, location information, and so on.

The combined raw dataset has over 400k records and 27 attributes. The recruiting data was captured since 2016 thus at the time of this research, we have 3 years of data. Due to ethical considerations for this project, we did not retrieve the candidate information, but only obtained the anonymous-aggregated information such as the count of clicks, applies, interviews, offers, hires etc. Table 1 illustrates the raw data attributes with respect to their functional relationship. i.e. The "Date of Activity" is the date that the "Candidate" performed an activity.

Table 1. Raw Data Attributes

Functional Relationship	Raw Data Attribute
Requisition/Job Posting	Requisition ID
	Job title
	Functional area
	City
	State
	Country
	Education Level
	Experience Level
	Location
Candidate	Date of activity
	Source Engine
	Visit
	Subscribes
	Apply Start
	Apply Complete
	Qualified
	Interviews
	Offers
	Hires
Cost Per Hire	Source Cost
	Cost per visit
	Cost per hire

Overall, the RMK, ATS, and job posting data provide a good overview of the journey from posting a job to hiring a candidate.

4   Methods

We condensed the final dataset down to four main attributes for ease of interpretation and as a proof of concept to expand to the entire dataset. Our final attributes consisted of Date, Visits, Source Engine, and Role.

The date attribute was taken from the Applicant Tracking system and represents the date at which the Visits were tracked. The Visits correspond to how many clicks a Job Posting of type "Role" and from "Source Engine" received on a date. The Source Engine was taken from the Recruiting Marketing system and corresponds to where the potential applicant came from. For our analysis, this was limited to LinkedIn Job

6

Postings. The Role was created using a subject matter expert to classify different jobs into Analyst, Engineer, Developer, Management, and IT jobs based on the Job Title. For our analysis, we limited this to solely the Analyst positions as those appeared to have the most accurate classification in terms of similarity of responsibilities.

To create the merged dataset, we obtained a drilldown version of the RMK and ATS datasets that consisted of Date, Requisition ID, Job Title, and Source Engine. The Visit counts were automatically aggregated to match this drilldown such that the counts of this visit corresponded to the Date, Requisition ID, Job Title, and Source Engine collectively. Using the Requisition ID as a link, we were able to augment this dataset with the internal Job Posting Data.

In the process of linking these datasets, we encountered missing values with respect to the requisition id. The requisition id is an identifier for a specific job posting based on the job title and functional area. In this case, a missing value signifies that this data corresponds to the running of a marketing campaign. In these campaigns, the primary goal is to expand the brand and increase the company's network in terms of a talent community database. These are not tied to any specific job. Thus, when encountered for our analysis, these data points were removed as they were irrelevant.

After the linking was completed, the dataset was sent to a subject matter expert to classify the job postings into separate categories or "Roles" based on the job title. This was an important step in the data cleaning process as we intended to do a time series analysis to predict how a job would fair if it were posted on LinkedIn. A typical posting is left on LinkedIn for a minimum of 2 weeks and a maximum of roughly 3-6 months. This means that each Requisition ID by itself will not have enough data or may be up to date for predicting the next week's visits. On top of that, the Requisition ID is unique for each job posting. So, in order to predict the visits for a new posting, some sort of clustering mechanism was required and in this case we chose an SME.

With the Role appended column, we reduced our dataset based on Analyst Roles and LinkedIn Job Postings as the Source Engine. We then mean-aggregated the number of visits based on the date such that we get the average number of visits per day an Analyst Role would receive from a LinkedIn Job Posting. The choice of the mean was important as there was no guarantee that we would have an equal number of Analyst postings each day. This would also eliminate the bias obtained from the demand the company had for a job during different times of the year. This also worked well for the business case of the model as it would be used to predict the visits for a single job posting rather than a group of job postings. With that we had a cleaned dataset of daily visits for Analyst Job postings on LinkedIn.

The time series models were trained on the first two-thirds of this cleaned dataset and tested on the last one-third. We used the python statsmodels library as our implementation of the AR, MA, and ARMA models.

5   Time Series

We utilized a time series approach for the analysis. Often, there are certain jobs that get posted throughout the year every year. There may be seasonal dependencies on the number of applications or visits that this type of posting receives which lends itself towards a time series approach. In another business scenario, a recruiter may have a need to find a number of applicants quickly for an urgent position. In this case, a model that provides a week to week forecast of how many applicants or visits a particular posting can expect can help the recruiter optimize their budget spending to fulfil their short-term goals. One last reason to pursue a time series approach is that a good classical time series model can serve as a baseline for future research [7].

A time series is a sequence of data that is ordered by a date or a timestamp. Generally, time series data has an equal time step between the points, i.e. daily, weekly, monthly, etc. time steps. The goal of a time series analysis is to extract meaningful statistics and other characteristics from Time Series data and potentially forecast values of interest. Exploring the statistics and characteristics of the time series data enables a more rigorous selection of a time series model [8]. The model is then generally used to make forecasts using previously observed values [9].

There are many approaches to time series analysis. A classical time series approach involves algorithms such as Autoregression (AR), Moving Average (MA), Autoregressive Moving Average (ARMA), Auto-regressive Integrated Moving Average (ARIMA) and more [10]. Non-classical methods involve the use of machine learning methods, i.e. neural networks, random forest, SVM, etc. Makridakis, Spiliotis, and Assimakopoulos investigate a comparison of traditional time series methods with that of the machine learning methods and found for simple univariate time series, the machine learning methods generally underperformed compared to the traditional statistical approaches [7]. With that, they suggest that classical time series approaches make a good baseline for moving forward with machine learning methods that hope to improve on the classical methods.

The classical time series models for time series data encapsulate several forms and are best used on different stochastic processes. The main models exhibited in literature are the MA and AR models. The AR model is a linear regression using a number (p) of past observations, an error term, and a constant term to predict the future value. The MA model is another linear regression that uses a number (q) of past error terms as the explanatory variables. These two methods can also be combined into the ARMA(p, q) method [8]. The AR, MA, and ARMA methods require a stationary dataset which means that the mean, variance, and autocorrelation structure do not change [11]. To observe the autocorrelation structure of the data, the autocorrelation function (ACF) and partial autocorrelation function (PACF) are generally used.

The ACF and PACF both measure the correlation between the current and past values of a series. The plot of these correlations indicate which "lags" of the response variable may be important in the AR, MA, and ARMA models [12]. The difference between the ACF and PACF model is that the ACF includes "direct and indirect independence information" when calculating the autocorrelation and the PACF only considers the dependence information [13].

8

Time series analysis can be applied to real-valued, continuous data, discrete numeric data, or discrete symbolic data. The power of time series can be applied to several real life examples. For instance, US Housing data. Data was collected for houses sold in US from 1963 to 2016. Using time series models one can predict how housing sales volume has changed over the years along with seasonality [14]. Another example involves a large food distribution company in Australia and New Zealand. The company used the time series model to forecast demand for products across numerous warehouses. Using these forecasts they were able to efficiently replenish warehouse stock. They use hybrid time series models consisting of both heuristics and classical techniques [15].

6   Preliminary Analysis

For the preliminary analysis, we created 3 separate datasets from our cleaned data set which sum-aggregated the visits into daily, weekly, and monthly time steps. This was done to explore the various autocorrelations between different levels of time step. We first fitted an ordinary least squares model to these datasets and viewed the residuals fitted with a LOESS smoothing function as shown in Figure 1. In Figure 1 we see that autocorrelation appears to be more prevalent for the weekly and monthly time steps vs the daily. The daily data's LOESS line is flatter compared to the weekly, and monthly datasets. However, none of the datasets appear to show a significant upwards or downwards trend which suggests either stationarity or a linear trend for our dataset. Most importantly, we find that there appears to be an up and down pattern in the residuals. This pattern suggests that data points are not independent from each other and that the weekly visits exhibit autocorrelation.

We also note that the first and last points of the weekly and monthly datasets appear to be outliers. The first data point contains only 1 date. We defined weeks to start on Mondays. Since this is the first data point and the week's data did not start on a Monday, we will drop it. We also decided to drop the last data point as well. Since we would be predicting the next month's or weeks' worth of visits, we consider that last point to represent that "next week's" data. Therefore, it should not be included in our models as we would need the entire week or month to update the model.
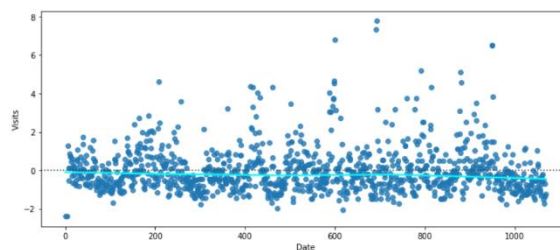

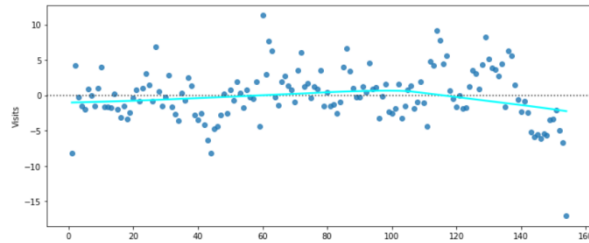**Fig. 1.** Residuals from an OLS model with a LOESS line over Visits for all Job Roles at Daily

Fig. 2. Residuals from an OLS model with a LOESS line over Visits for all Job Roles at Weekly timesteps. The residual plots suggest stationarity or a linear trend in the dataset. A pattern is prevalent in the residuals which suggest autocorrelation and there may also be some seasonal component.
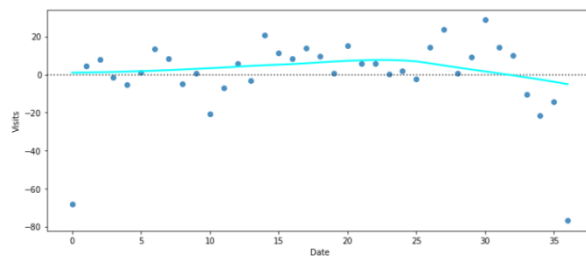

Fig. 3. Residuals from an OLS model with a LOESS line over Visits for all Job Roles at Monthly timesteps. The residual plots suggest stationarity or a linear trend in the dataset. A pattern is prevalent in the residuals which suggest autocorrelation and there may also be some seasonal component.

As a daily model is not necessary for the business case and we see evidence for autocorrelation within the weekly and monthly datasets, we will exclude the daily data from rest of the analysis.

Next, we investigate the additive seasonal decomposition of our weekly and monthly datasets (Fig. 2 & 3). We see no significant seasonal component and that the data is fully explained by the trend. Based on the Observed plots, we have a stationary data set, which meets an assumption that the classical Time Series techniques require (AR and MA).
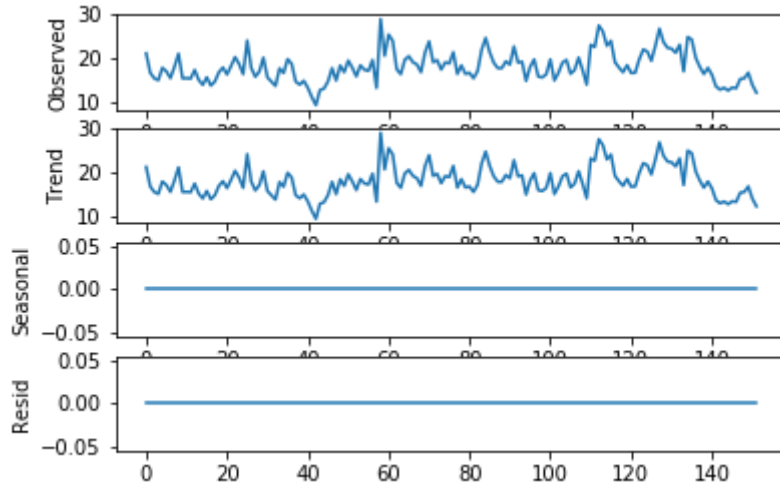
10

Fig. 4. Additive seasonal decomposition of the weekly data without first and last outlier. No seasonal or residual components were detected. The data is completely described by the trend.
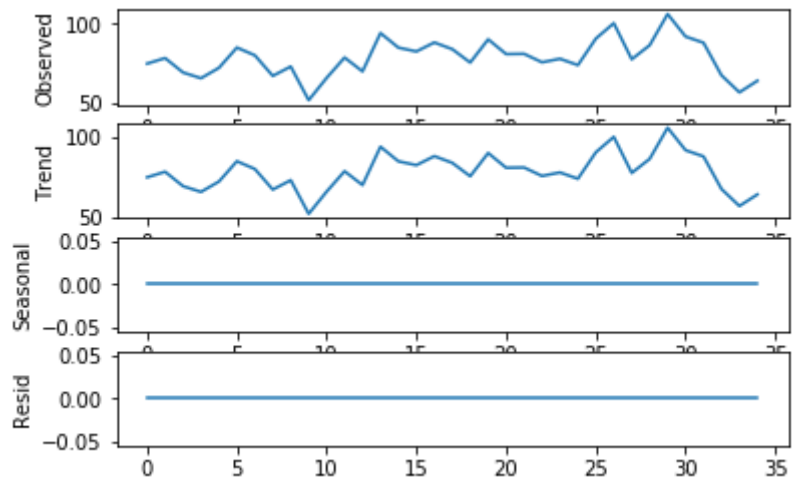

Fig. 5. Additive seasonal decomposition of the monthly data without first and last outlier. No seasonal or residual components were detected. The data is completely described by the trend.

Going further into the Weekly and Monthly time steps to investigate the autocorrelation, we examine the ACF and PACF plots in Figure 3. The ACF plot for the weekly dataset suggests that the responses are correlated with 1 to 4 lagged values whereas the PACF plot suggests lags at 1, 2, 19, 32, 35, 47, 52, 54, 55, and 59. As we are interested in simpler models and the ACF plot recommends 1-4 lags, we will consider the 1-2 lags for our models. It is also interesting to note that the 52-week lag shows

significance. The 52-week lag corresponds to the prior year's data. This matches our intuition about looking at what happened last year to determine what might happen this year. Therefore, we will be testing models with 1, 2, and 52-week lags for the weekly dataset.

The monthly ACF plot only shows a significant lag at 1. This is reinforced by the Durbin-Watson Statistic value of 0.886. This value is less than 2 which also suggests that the autocorrelation lag 1 value is significant. Interestingly, the PACF plot also only shows a significant 1 lag value. We had expected a significance at month 12 as well, as this would again correspond to last month's data. However due to having only 3 years of data, determining a 12-month lag only had 24 points to choose from. As a weekly-dataset is more fitting to our business case and the monthly data is a bit more limited by only having 3 years of data, we will continue evaluating our models with only the weekly dataset.
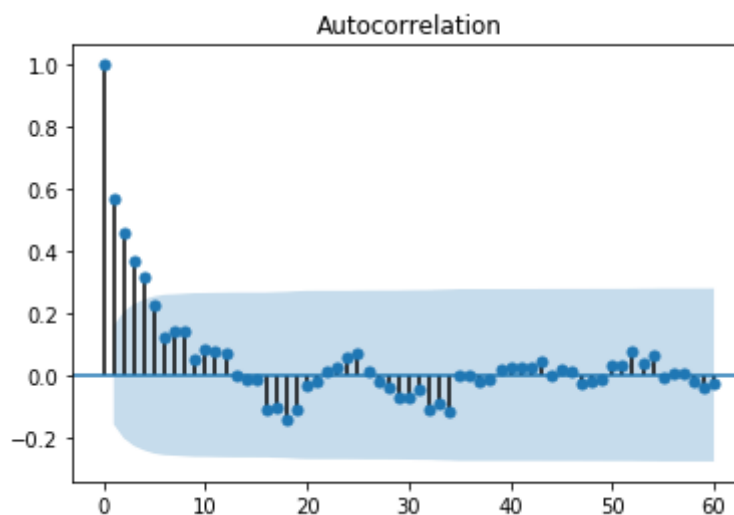

Fig. 6. ACF plot for weekly time steps. Durbin-Watson statistic for lag 1 weekly data is 0.835
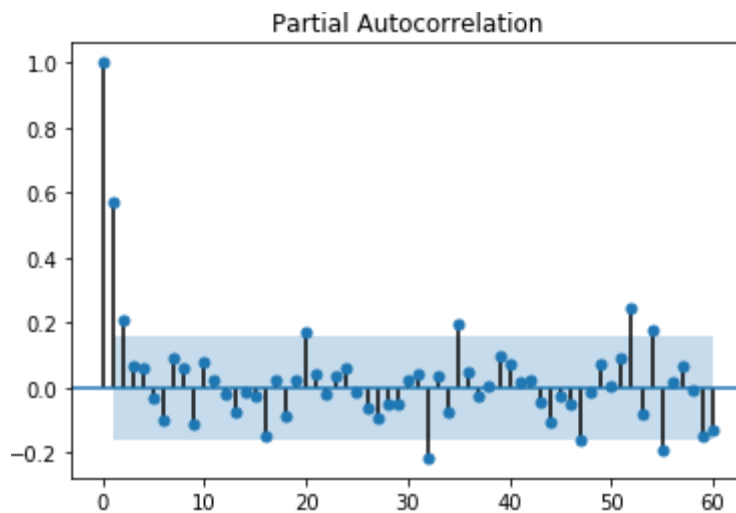
Fig. 7. PACF plot for weekly time steps. Durbin-Watson statistic for lag 1 weekly data is 0.835.
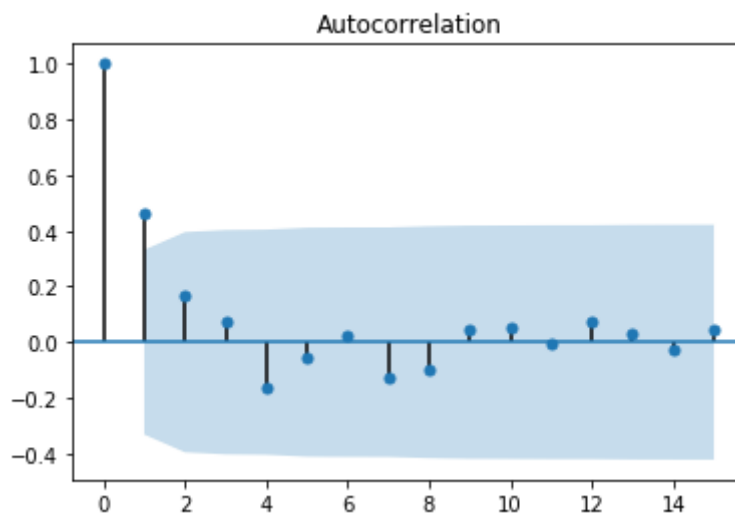

Fig. 8. ACF plot for monthly time steps. Durbin-Watson statistic for lag 1 monthly data is 0.886.
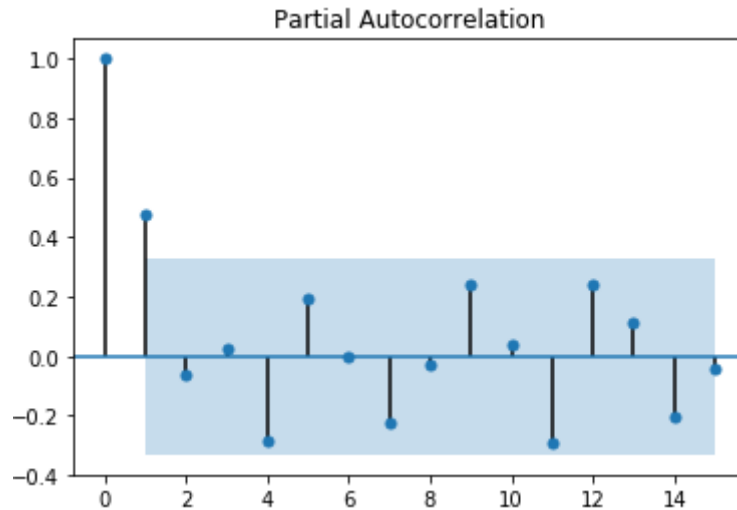
Fig. 9. PACF plot for monthly time steps. Durbin-Watson statistic for lag 1 monthly data is 0.886.

7   Modeling and Evaluation

We compared several models with the weekly timestep dataset: MA(1), AR (1), AR(2), AR(52), ARIMA(1,0,1), and a baseline persistence model which carries forward the last response value. We split the data into Train and Test using the first two-thirds of the data as a training set and the last one-third as a test set. To get the test predictions, we predicted one week's value at a time. We retrained the model using the last prediction's actual test point to predict the next and so on. To compare the models we considered the AIC and MSE metrics. Each prediction is for one week in advance. We use the mean squared error and the Alkali Information Criterion to compare the models. The MSE provides us information of how accurate the model is, whereas the AIC will also penalize more complicated models. Figures 10-15 show the predicted versus actual values for each of our models.
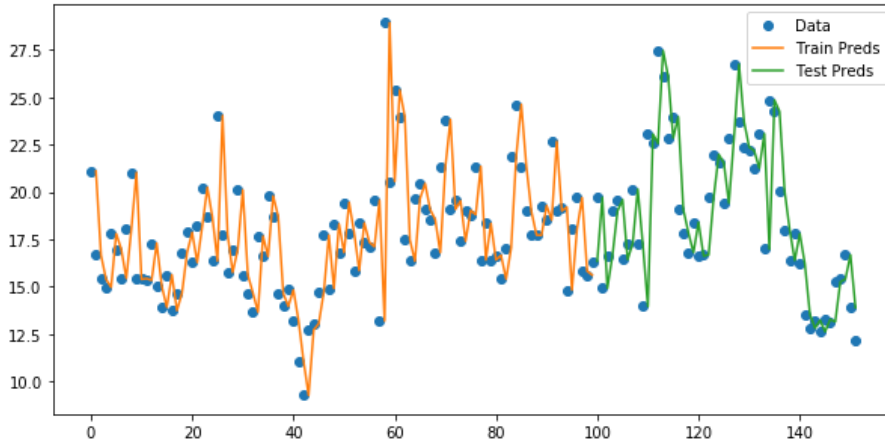
14

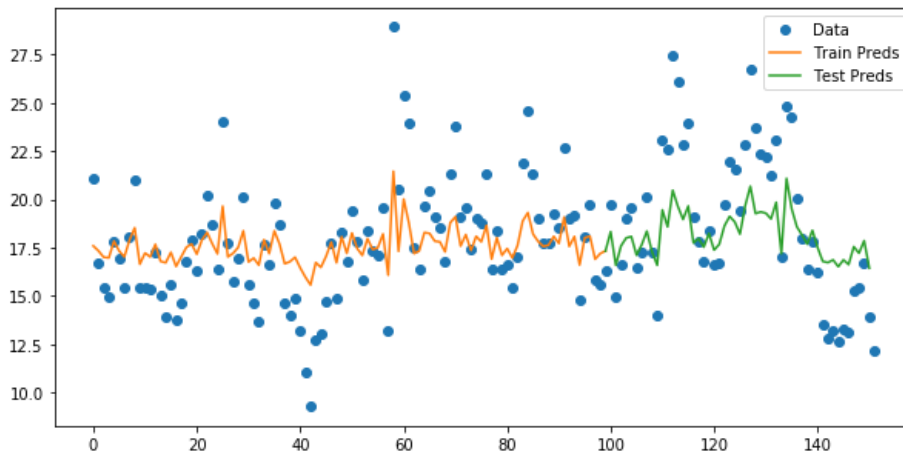**Fig. 10.** Plot for Baseline Persistence model: For weekly data.

Fig. 11. Plot for Moving Average: MA(1) for weekly data with a 10 week forecast.

Fig. 12. Plot for Auto Regression: AR(1) for weekly data with a 10 week forecast.

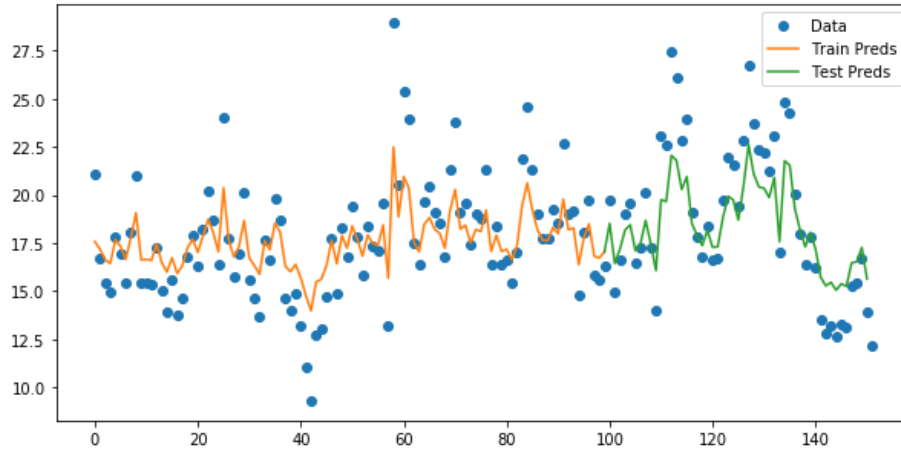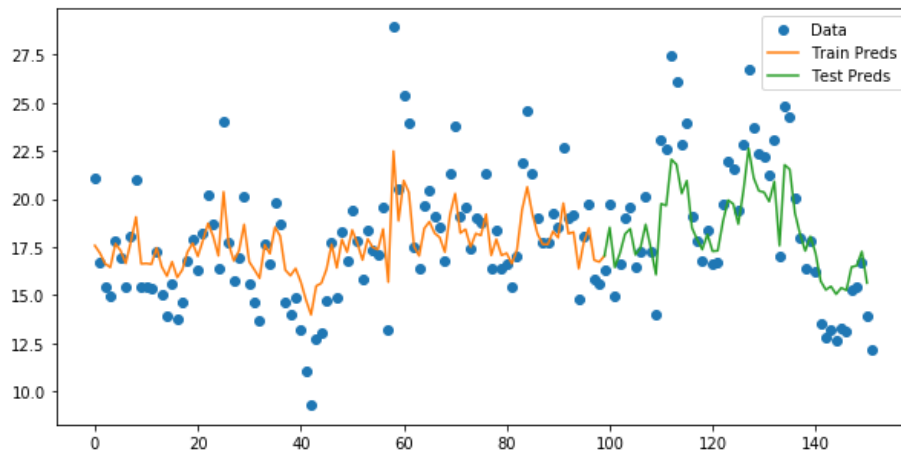**Fig. 13.** Plot for Auto Regression: AR(2) for weekly data with a 10 week forecast.
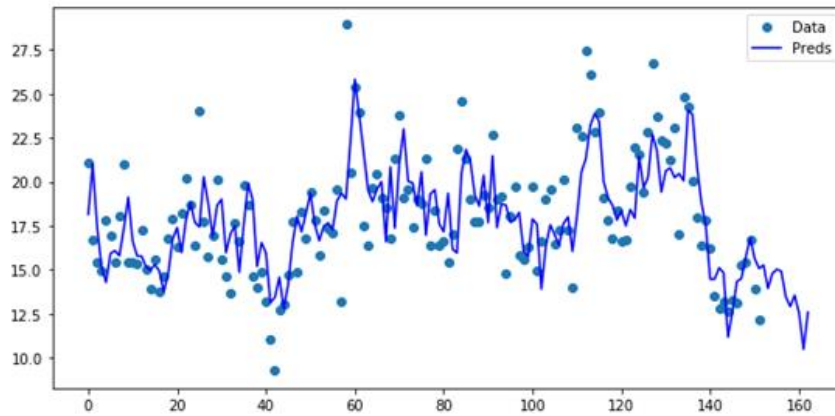
16

Fig. 14. Plot for Auto Regression: AR(52) for weekly data with a 10 week forecast. Note, we required all the data in order for the 52 lag parameters to converge. Therefore, we have no test error for this model.
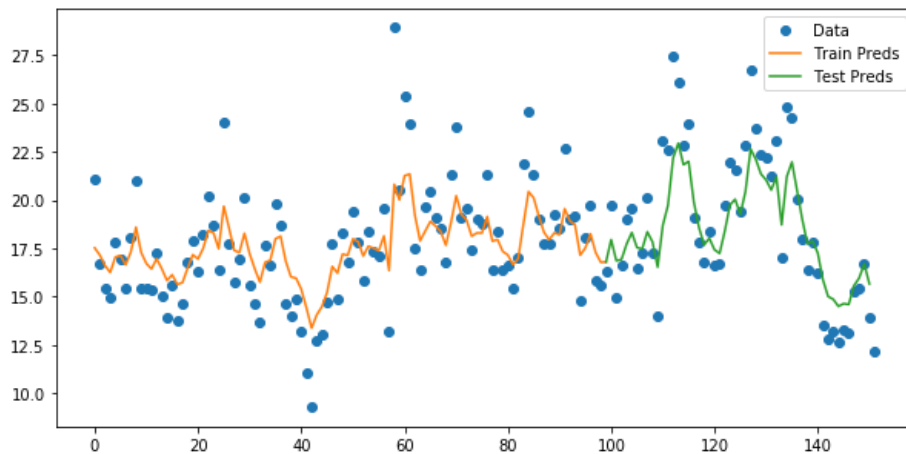

Fig. 15. Plot of Autoregressive Moving Average: ARIMA(1,0,1) model for weekly data with a 10 week forecast.

Table 2. Model Comparison Results

Model	Train MSE	Test MSE	AIC
MA(1)	8.139	11.528	775.771
AR(1)	7.558	8.719	2.101
AR(2)	7.207	8.421	2.066
AR(52)	6.054	--	2.459
ARIMA(1,0,1)	7.199	8.325	735.479
Persistence	10.707	8.642	--

From the results shown in Table 2, the ARIMA(1,0,1) model performed the best in terms of the Test MSE with a value of 8.325. However, its AIC value is far too high compared to our pure AR models (735 compared to 2). The AR(1) and AR (2) models are quite similar to each other with Test MSE scores of 8.719 and 8.421 and AIC scores of 2.101 and 2.066 respectively. The MA(1) model does not compare to the rest with a high Test MSE of 11.528. The AR(52) model looks promising in terms of the train MSE, however due to modelling convergence issues we were unable to obtain a test MSE. Thus, the AR(2) model exhibits the best behavior with respect to both the MSE and AIC values. This aligns with our preliminary analysis where we saw 2 significant lags in our PACF plots. However, none of the models are significantly better than the persistence model in terms of practical significance. Our best Test MSE model is off on average by 2.89 visits whereas the persistence model is off by 2.939 visits. Overall, the Persistence model may be the best in terms of practicality for training time, ease of implementation, and overall accuracy for this dataset.

8   Ethical Considerations

Ethics play a very important role during the recruiting process. While there are laws in place that protect certain individuals during the hiring process--as well as help maintain a professional, ethical standard when hiring employees--HR professionals are often faced with dilemmas that extend beyond these principles. Some common ethical dilemmas in hiring can include:

18

- Placing misleading advertisements for jobs
- Misrepresenting the requirements of a particular position.
- Considerations when collecting data

8.1   Placing misleading advertisements for jobs

Job ads for positions that actually differ from what is being advertised should not be placed. Job scams and misleading job ads are ways that people try to take advantage of people looking for work. Tremendous pressure from the competitors could sometimes lead recruiters to make unethical decisions. To tolerate the pressure, some Recruiters might mislead applicants about openings or about their chances of getting a job. They might post fake job descriptions or fabricate a relationship with an employer. Some recruiters misuse applicants' personal information.

Most job ads are real and recruiters behave ethically knowing their job and company reputation are at stake. There is also a discussion to stem the unethical behavior recruiting field itself would need licensing and code of ethics. The Association of Talent Acquisition Professionals recently rolled out a set of guidelines for recruiting integrity, as has the National Association of Executive Recruiters and the American Staffing Association. The Brussels-based World Employment Confederation has a code of conduct [16, 17]. The consequences of these false representations can be devastating. Unlike fraud, misrepresentation of requirements is where an employer exaggerates the responsibilities of a position.

8.2   Ethical considerations when collecting data

Various ethical issues should be considered when planning for data collection, as it might cost either the person collecting the data or the company providing the data. Project outcome is critical for data collection. For this project, the question could be if the predictions would increase the visits / applies for the job posting.

Based on the data, the projects draw some conclusions / methods for implementing the conclusions which might result in additional resources, budget and time. Bad data or poorly collected data could result in incorrect conclusions which could also result in inefficiencies and waste of resources, budget and time. [18]
Therefore, it is important to plan what data will be collected from the beginning of the planning process. We should ask if the data will "truly be needed, be disseminated widely, be used to implement or revise a program, and use the least intrusive and costly data collection method possible." [19]

Irrespective of the type of data we are collecting, we need to get the approval of the company, stakeholders, or community where the data is being collected. [19]

9   Conclusions and Additional Research

In conclusion, we found the AR(2) model to be the best to describe our dataset in terms of the AIC and MSE. Our model can forecast the next weeks data with an aver-

age squared error of approximately 8.421 squared visits. This model was run over a limited dataset of Analyst job postings on LinkedIn. However, in a practical sense this model does not perform much better than the naïve persistence model for this dataset. It is important to note that there are many different source engines, and job roles besides LinkedIn and Analyst in which we analyzed. This practical significance may not be true for other engines and role categories. Extending these models in practice will require the involvement of other source engine data (Indeed, Monster, etc.) and will require the data and model to be updated on a weekly basis. This model provides insight into how many visits a recruiter can expect if he/she posts a job to LinkedIn. This insight, given the model is extended to multiple job boards, can then help determine where to post jobs, based on maximizing the number of visits, and reduce overall budget spending if it is found only a subset of available job boards are needed.

Often, knowing the number of applicants that one will receive for a job posting is more important than knowing the number of visits a job posting will receive. With the current model this will need to be calculated by some "rule of thumb" percentage that converts visits to applicants. However, this leads to future research in extending our model to forecast the number of applicants. This would involve the inclusion of a multivariate time series model as we expect the number of applicants to be dependent not only on the lagged number of visits and applicants, but also the forecasted visits. We expect there to be a more practical difference between the persistence and trained models in this scenario. Also, in this manner, a "rule of thumb" would no longer be necessary.

A way to potentially improve the model is to categorize postings using NLP on the job description. To divide the data into appropriate roles, we used a domain expert to categorize jobs based on title. The dataset contains plenty of job titles with the word Analyst, but the meaning can vary significantly across different business areas. For Ex: an Analyst could either be an IT analyst or an Investment analyst. Both have analyst in the title, however there is a huge variation in terms of job duties, pay structure, and even the recruitment process. It may also be desired to automate this task of categorizing job postings into a specific role or generic category. NLP has this potential application to cluster similar roles based on the job description, requirements, title, and so on.

Lastly another application of our model may include Recruiting Budget Forecasting. Based on the hiring from previous years and forecasting on the number of visits through this model, we can set a more accurate budget for recruiting marketing in the year ahead and the resources need to fulfill the job. This would involve extending our forecast to a yearly basis and potentially using a more aggregated dataset such as the monthly data.

References

1. "campaign," [Online]. Available: https://www.topechelon.com/blog/marketing/successful-recruitment-campaign-ideas-strategy/.

20

2. SHRM. [Online]. Available: https://www.shrm.org/resourcesandtools/tools-and-samples/hr-qa/pages/ethicaldilemmashrprofessionalsface.aspx.

3. L. Ethington, "How data science is changing the way we are recruiting," Allegis Global Solutions, 4 April 2017. [Online]. Available: http://blog.allegisglobalsolutions.com/how-data-science-is-changing-the-way-we-are-recruiting. [Accessed 10 3 2019].

4. "Data Science vs Analytics: The Expertise Your Marketing Needs Most Now," Market Pro, [Online]. Available: https://marketproinc.com/hiring-advice-2/2015/12/marketing-analytics-recruiters-data-scientists. [Accessed 10 3 2019].

5. "Job Board, Hiring Success Glossary," Smart Recruiters, [Online]. Available: https://www.smartrecruiters.com/resources/glossary/job-board/. [Accessed 12 3 2019].

6. S. Makridakis, E. Spiliotis and V. Assimakopoulos, "Statistical and Machine Learning forecasting methods: Concerns and ways forward," PLoS ONE, vol. e0194889, no. https://doi.org/10.1371/journal.pone.0194889, p. 13, 2018.

7. R. Adhikari and R. K. Agrawal, "Time Series Forecasting Using Stochastic Models," in An Introductory Study on Time Series Modeling and Forecasting, Arxiv.org, pp. 18-24.

8. "Time Series," [Online]. Available: https://en.wikipedia.org/wiki/Time_series. [Accessed 1 3 2019].

9. J. Brownlee, "11 Classical Time Series Forecasting Methods in Python," Machine Learning Mastery, 6 Aug 2016. [Online]. Available: https://machinelearningmastery.com/time-series-forecasting-methods-in-python-cheat-sheet/. [Accessed 31 3 2019].

10. "Stationarity," Engineering Statistics Handbook, [Online]. Available: https://www.itl.nist.gov/div898/handbook/pmc/section4/pmc442.htm. [Accessed 31 3 2019].

11. "Autocorrelation and partial autocorrelation functions," IBM Knowledge Center, [Online]. Available: https://www.ibm.com/support/knowledgecenter/SS3RA7_18.2.0/modeler_mainhelp_client_ddita/components/dt/timeseries_acf_pacf.html. [Accessed 31 3 2019].

12. J. Brownlee, "A Gentle Introduction to Autocorrelation and Partial Autocorrelation," Machine Learning Mastery, 6 Feb 2017. [Online]. Available: https://machinelearningmastery.com/gentle-introduction-autocorrelation-partial-autocorrelation/. [Accessed 31 3 2019].

13. "Housing Data," [Online]. Available: https://dzone.com/articles/investigating-real-world-data-with-time-series.

14. N. Wagner, Z. Michalewicz, S. Schellenberg, C. Chiriac and A. Mohais, "Intelligent techniques for forecasting multiple time series in real-qorld systems," International Journal of Intelligent Computing and Cybernetics, vol. 4, no. 3, pp. 284 - 310, 2011.

15. "Recruiting Integrity," [Online]. Available: https://atapglobal.org/ethics/.

16. "Capital Group Career Site," [Online]. Available: jobs.capitalgroup.com.

17. "What is the ethical ways to collect data," Research Gate, 10 Oct 2017. [Online]. Available: https://www.researchgate.net/post/What_is_the_ethical_ways_to_collect_data. [Accessed 31 3 2019].

18. S. Bates, "Code of Ethics," 9 Jan 2019. [Online]. Available: https://www.shrm.org/resourcesandtools/hr-topics/talent-acquisition/pages/do-recruiters-need-code-of-ethics.aspx. [Accessed 31 3 2019].

19. "Job Scams," [Online]. Available: https://www.youthcentral.vic.gov.au/jobs-and-careers/how-to-find-a-job/job-scams-and-misleading-job-ads.

20. "Hiring Cost," [Online]. Available: https://www.shrm.org/about-shrm/press-room/press-releases/pages/human-capital-benchmarking-report.aspx.

21. "An End-to-End Project on Time Series Analysis and Forecasting with Python," Towards Data Science, 8 Jul 2018. [Online]. Available: https://towardsdatascience.com/an-end-to-end-project-on-time-series-analysis-and-forecasting-with-python-4835e6bf050b. [Accessed 20 3 2019].
22. "Mathworks," [Online]. Available: https://www.mathworks.com/help/econ/multivariate-time-series-data-structures.html. [Accessed 31 3 2019].

Appendix

Field Attributes:

This dataset has 16 fields which are tied to attract, engage and hiring of top talent.
The recruiting Marketing data contains the attract and connect component of candidate journey. These consist of fields visits, subscribes, Apply starts. The applicant tracking system contains the engage and hire component of candidate journey. This has fields like Apply complete, Qualified, Interviews, Offer, hire etc. The cost related fields are the Source cost and cost per hire.

Source type is the origin of how a candidate has viewed, joined or applied for the job. This information is key for any company and will help to put strategy around the marketing campaigns. There are several sources from where the information can be collected. A brief overview of each source type is listed below.

Alumni:
Alumni specific social networks or any online source used for alumni recruiting.

ATS Prospecting:
Source Engines including Xpten & E-Rec.

Banner Campaign :
Custom banner campaigns across targeted advertising networks.

Blogs:
All large blog networks are categorized under the "blog" source type. No tracking codes are needed for standard blog traffic that is driven through the RMK platform.

Career Site:
Career site traffic includes links that point to an RMK landing page from your career site.

22

Career Site Service:
RMK "widgets" that are deployed on your career site or anywhere online. Includes search Box, Map Widget, E-mail subscribe, Email Subscribe Link, and Category Group Menu.

Custom Campaigns:
The custom campaign type is used for vanity URLs or other campaigns that can't be categorized by other sources types. The campaign names are used to differentiate between multiple custom campaigns.

Direct:
A direct visitor will occur when no referrer is present during a candidate visit. Reasons for no referrer being captured include :

- Link from https to http
- Blocked by security settings
- Browser Settings
- Firewall(user slide)
- URL entered directly in browser
- Bookmark
- Link not from a web browser
- Non-web email client(Outlook)
- Instant Message

When a visitor is categorized as "direct" the RMK system was not able to capture any type of referral source data for the visit. Visitors cannot be manually tagged with a "direct" source.

Email:
All Talent Community emails sent from the RMK platform are automatically tagged with accurate URL source tracking.

- Email Admin : Talent Community email to change user settings
- Email Campaign : Client runs email campaign and specifies source with code.
- Email Signature : Client recruiter includes link in email for candidate to respond.
- Email Subscription : Talent Community email alerts candidates clicking through the job code.
- Mobile Apply : Mobile apply email alerts candidates clicking through to view jobs.
- Online Email : Candidate clicking to link from their online email (gmail etc)

Traffic will be associated with "online email" when an URL is emailed to a candidate without proper URL tracking parameters.

Employee Referral marketing:
Traffic will be associated with the "Employee Referral Marketing" when candidates click on links generated by the Employee Referral Marketing program.

eNewspaper:
Newspapers that have an online presence are categorized as "eNewspaper". Many of these sources may also be news aggregators that have advertising areas. Some of these eNewspapers also have job board sections.

Events:
The "events" source type is used for conferences, job fairs, or any offline networking event.

Industry Groups:
Associations or industry online communities where the primary purpose is professional development and networking. These sites also include a "Jobs" section on their sites as a service to their members – as opposed to a job board where the primary purpose of the site is job postings or content

Imported Talent Community Members:
They are candidates who is client talent community but who do not have any referring information.

Job Aggregators:
Search engines specifically for jobs. Job aggregators scrape and capture job listings from a board list of employers and other major sources, and compile them into one search system. As a general rule, clients don't post jobs individually to these sites. Job aggregators typically have "organic" free sections as well as a sponsored "pay per click" section.

Job Board-Major:
Major job boards include Monster, CareerBuilder, Dice, The Ladders and Hotjobs. Some boards are further delineated to indicate "products".

Job Board – Niche:
The primary purpose of Niche Job Boards is job postings and content . The Niche job board source type is compromised of any job boards not included in the Job Board – Major source type listed above.

Jobs2Web:
The Jobs2web source type includes any traffic that is generated from jobs2web.com. Jobs2web may highlight a client site as part of a case study or other marketing initiative.

Media:
Offline media including billboards should be associated with the "Media" source type.

24

Mobile:
Any traffic that is generated through RMK mobile websites will be tagged as "Mobile".