

## A PageRank-Based Algorithm for Real Estate Agent Ranking

Kai-Wen Lee, Graduate Student, Computer Science,  
University of Miami, [kx1864@miami.edu](mailto:kx1864@miami.edu)

Dr. Vladimir Bugera, Adjunct Faculty, Journalism and Media Management,  
University of Miami, [vbugera@miami.edu](mailto:vbugera@miami.edu)

Prof. Sergiy Butenko, Professor, Industrial and Systems Engineering,  
Texas A&M University, [butenko@tamu.edu](mailto:butenko@tamu.edu)

Dr. Aidar Vafin, Director,  
Big Data Realty, [aidar@bigdatarealty.com](mailto:aidar@bigdatarealty.com)

### 1. Abstract

We applied an analytical method, PageRank, to rank the licensed real estate agents based upon the real estate offers where the agent represents a seller or a buyer of a residential real estate property, and the offers resulted in the property sales with the corresponding parties. We analyzed the resulting ranking by studying patterns and common characteristics for agents with high and low scores. Our dataset is based on the listing data obtained from Miami Association of Realtors. We used Python programming for data pre-processing and algorithmic ranking.

### 2. Introduction and Background

For many years, most of us in one way or another has been faced with the questions of choosing a realtor when selling and / or buying real estate. Even though, the answer to this question seems rather trivial, in fact, how to really understand, based on facts, and not on emotions, that one or another agent is better than the other? We will not surprise anyone by saying that each of us wants to work with the best realtor, and in the standard scenario, we seek the help of acquaintances and friends, look, and analyze different ratings and reviews. However, to what extent do people's reviews really reflect the effectiveness and efficiency of the realtor's services? Unfortunately, reviews do not have such parameters, and they most likely reflect the emotional part of this issue.

To answer the question posed, which literally sounds like this: “Which realtor should I choose and by what parameters do I understand that he suits me?”. We used big data analysis base on the data from the Miami Association of Realtors to analyze 25 objective parameters, including property type, zip code zone, listing price, sold price, relationship type, and many more. It is no secret that every seller (landlord) of a property wants to find a realtor who can sell (rent) for the highest price, and at the same time, every buyer wants to find a realtor who can help buy (rent) for the lowest price [1].

In this regard, we decided to objectively evaluate the performance of a realtor from a scientific point of view by applying PageRank algorithm [2]. The algorithm has been developed and used by Google to rank web pages in their search engine results. Since the original implementation, the algorithm has been adapted to many practical applications [3]. We designed dashboards using the interactive data analytics software Tableau Software. These dashboards serve to ensure that the client, in our case, the buyer or seller of a property, can easily find and view the parameters he needs to select the right realtor more accurately.

When developing a methodology for ranking realtors according to given criteria, we actively used PageRank network algorithms to explore the relevance of the nodes (participants) based on the links connecting them. The original algorithm is applied to a collection of hyperlinked documents (such as web pages from the World Wide Web) and assigns each of them a numerical value that measures its "importance" or "authority" among other documents. Generally speaking, the algorithm can be applied not only to web pages, but also to any set of objects interconnected by mutual links. To paraphrase in relation to the real estate market, we can say that realtors are like a "Web Page" that market participants link to and the more links to them, the more transactions they have and, accordingly, using the selected criteria, you can present an objective assessment of efficiency and rank them. In this regard, using our methodology, we came up with new approaches, such as Transaction-Based Model, Sales-Based model, Simplified Transaction Model for evaluating realtors, and the results obtained during our study are significant, meaningful, and practically applicable.

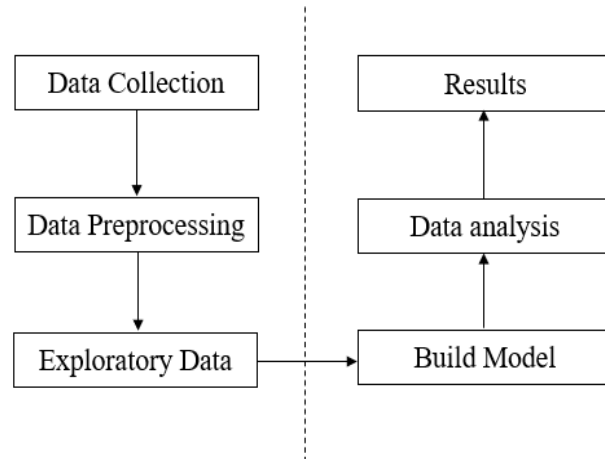
### 3. Methodology

#### 3.1 Page Rank Algorithm

PageRank is an algorithm developed and used by Google to rank web pages in their search engine results. Today, the methodology is one of the mostly recognized network algorithmic frameworks that helped to transform Internet Industry and changed how people access the information. Pagerank includes nodes and links. Node indicates the number of directed relationships that is measured the important for in-degree and out-degree. Links are used to connect each of nodes. PageRank is a network algorithm that can explore the relevance of the nodes (participants) based on the links connecting then. PageRank is used in many different fields, such as social networking, sports, and financial. Using PageRank can understand and calculate the influence of customers. Therefore, PageRank is widely used in analyzing broad influence over the network. When the value of PageRank is larger, it implies that the node has the higher influence, and it has the less likely to lose customers. Thus, development and implementation of the PageRank model can identify business opportunities and strategic changes. It is beneficial to companies' growth and business outcomes.

#### 3.2 Design methods and steps of models

In the design of the steps of models, it consists of selecting data, processing data, building models, and analyzing data. At the beginning, the study collected the data and did data preprocessing. After cleaning data, we designed three different models to compare which model is better for the ranking of real estate agents. In addition, we also design several dashboards to analyze the results of these models. (see Figure 1).



**Figure 1.** Design the steps of models

### 3.3 Data Preprocessing

We will use residential sales transaction data from Miami Association of Realtors. The data consisted of all residential listing (Condo / Townhomes / Single Family) from Miami Dade, Broward, and Palm Beach. We include both for-sale and for-rent transactions. We annualized the size of the rental transactions by multiplying monthly rent by 12. The dataset included 112,357 purchase and lease transactions occurred between January 1<sup>st</sup> 2021 and December 31<sup>st</sup> 2021. In total, 32,705 agents were involved in the selected transactions.

The original dataset had 25 attributes. We performed data preprocessing to remove duplicates, convert formats of several fields, and select only attributes needed for the analysis, such as agents' names and license numbers, type of property, sold price of the transaction, closing date, and others. After cleaning and exploring data, we selected attributes presented in the Table1.

Column name	Data type	Definition
MLSID	Character	Multiple Listing Services ID
RELTYPE	Character	Relationship Type
ADDRESSI	Character	Property Address for Display on Internet
BRK_NAME	Character	Broler Name
AGT_ID	Character	Agent ID
AGT_NAME	Character	Agent Name
PROPTYPE	Character	Property Type
LISTTYPE	Character	Lisitng Type
ORIGLISTPRICE	Numerical	Original Listing Price
LISTPRICE	Numerical	Listing Price
LIST_DT	Character	Listing Date
SOLDPRICE	Numerical	Sold Price
SOLD_DT	Character	Sold Date
ZIPCODE	Character	Zipcode

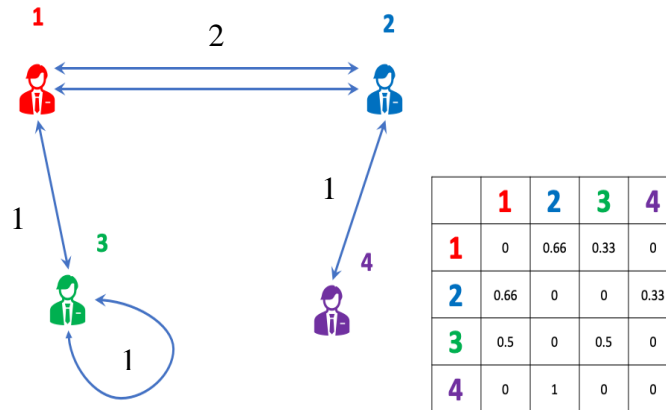
**Table 1.** Attributes of real estate

### 3.4 Build models

PageRank algorithm is used in various ranking applications where data can be modeled using network. It can help people easily find and retrieve appropriate information [2]. PageRank considers the relationship between nodes and their neighbors. The basic assumption is that it measures the quality and number of network connection to get importance of nodes. In our study we modelled agent relationships by an undirected graph, where each graph node corresponds to an agent, and each graph link corresponds to a sale transaction where agents are involved representing a buyer and a seller. Sometimes two agents may represent a seller, or two agents may represent a buyer, in which case the graph will have multiple links corresponding to all possible one-to-one combinations between buyer and seller agents.

We considered three separate models to address various objectives of the analysis. The first model is Transaction-Based model, where we focus on frequency and diversity of transactions but not on transaction amount. Transaction-Based model has the following basic assumptions:

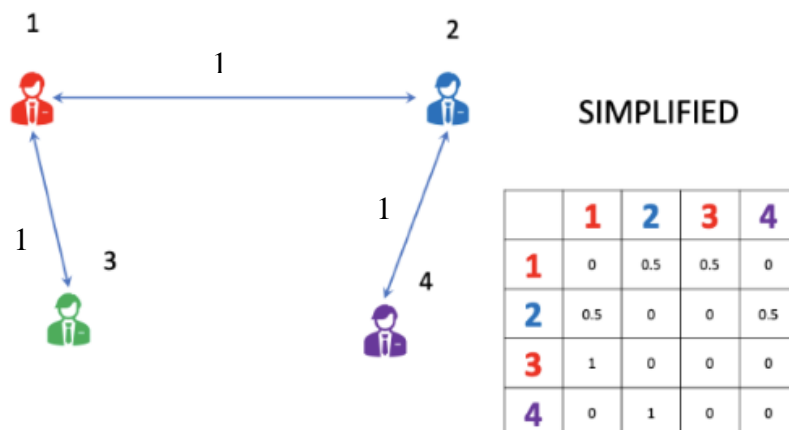
1. Each agent is represented by a node in the graph.
2. If two agents work on the same real estate transaction representing buyer and seller, then a link connects the corresponding nodes in the graph.
3. The graph is undirected (it does not matter which agent represents buyer or seller).
4. If an agent represents both buyer and seller, then there will be a loop link.
5. Weight of each link is the total number of transactions between the agents.



**Figure 2.** Transaction-Based Model

The second model is Simplified Transaction-Based model, where we ignore frequency of transactions between the same agents and remove “loop” links where the same agent represents both the seller and the buyer. The Simplified Transaction-Based model has the following basic assumptions:

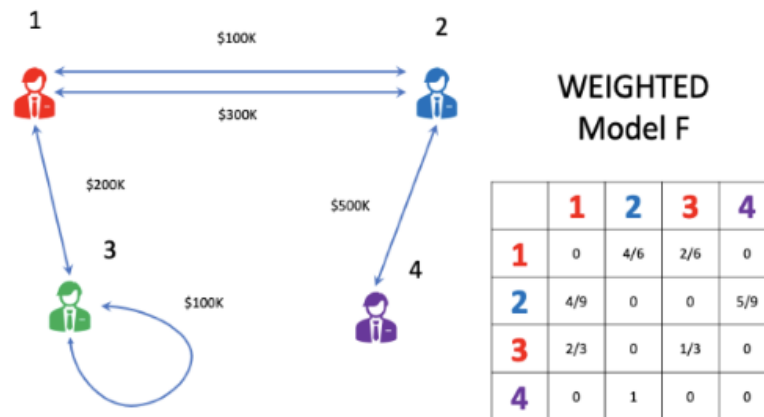
1. Each agent is represented by a node in the graph.
2. If two agents work on the same real estate transaction representing buyer and seller, then a link connects the corresponding nodes in the graph.
3. The graph is undirected (it does not matter which agent represents buyer or seller).
4. There are no loop links where an agent represents both buyer and seller
5. There is only one link If there are multiple transaction between two agents, the corresponding nodes will connect with multiple links. The weight of each link is 1.



**Figure 3.** Simplified Transaction-Based Model

The third model is the Spend-Based model. The prime motivation of the model is to evaluate agents not only by the strength of their networks but also by the size of their transactions. The weight of each link will correspond to the total transaction volume between the two agents. The Spend-Based model has the following basic assumptions:

1. Each agent is represented by a node in the graph.
2. If two agents work on the same real estate transaction representing buyer and seller, then a link connects the corresponding nodes in the graph.
3. The graph is undirected (it does not matter which agent represents buyer or seller).
4. If an agent represents both buyer and seller, then there will be a loop link.
5. Weight of each link is the total transaction volume between the agents.



**Figure 4.** The weight of Sales Based Model

### 3.5 Analysis Method

To analyze the ranking of the agents and performance of the models we designed several dashboards using interactive data analytics software Tableau. Each dashboard includes the overall agent sales statistics, the agent sales by zipcode, and the agent ranking.

These six worksheets are described as follows. In the worksheet of the sales of agent names in each zipcode, we found each agents sold and transact the number of houses. We used map to present the sales and transaction of agents to provide more intuitive information for real estate agents when they sold houses to customers. In addition, we can find the top 50 agents that have higher sales and transaction in the worksheet of weight of agent names and found why people buy houses in the related information of real estate agents table. Finally, we design the worksheet of weight of agent names in each step to understand to generate weights for each iteration.

Second, based on these worksheets, we designed three dashboards to make users clearly find and view all the most important metrics at a glance. (See Figure 5, Figure6, and Figure7).

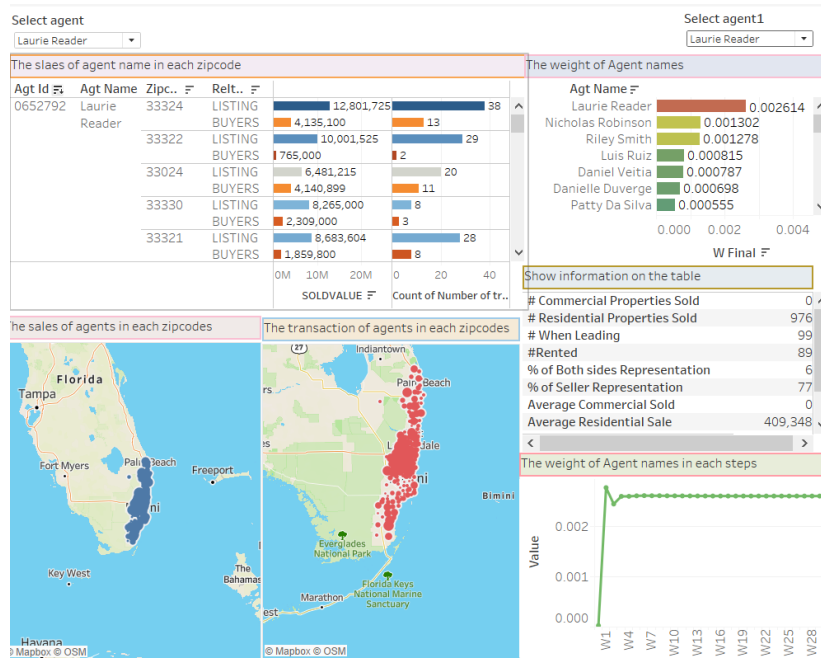


Figure 5. Transaction Based Model dashboard

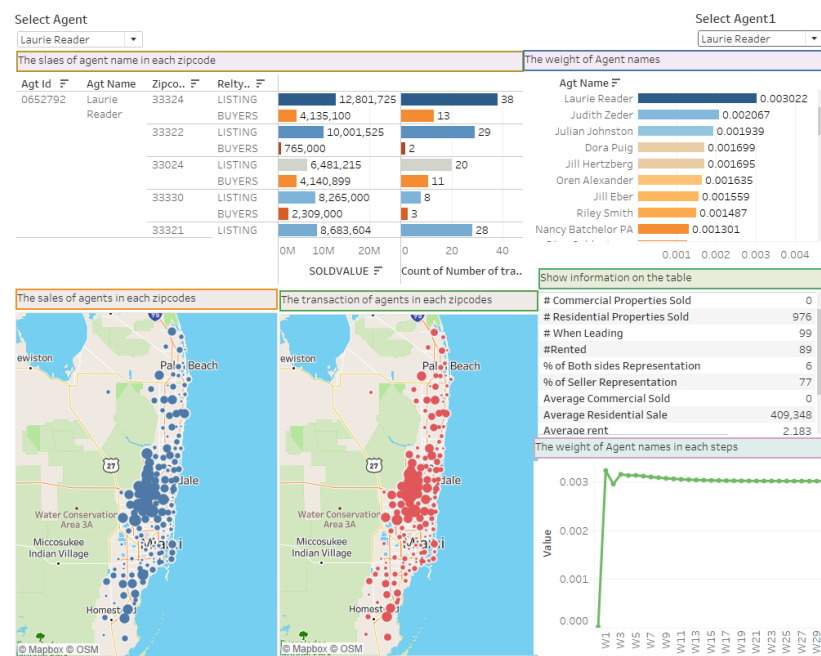
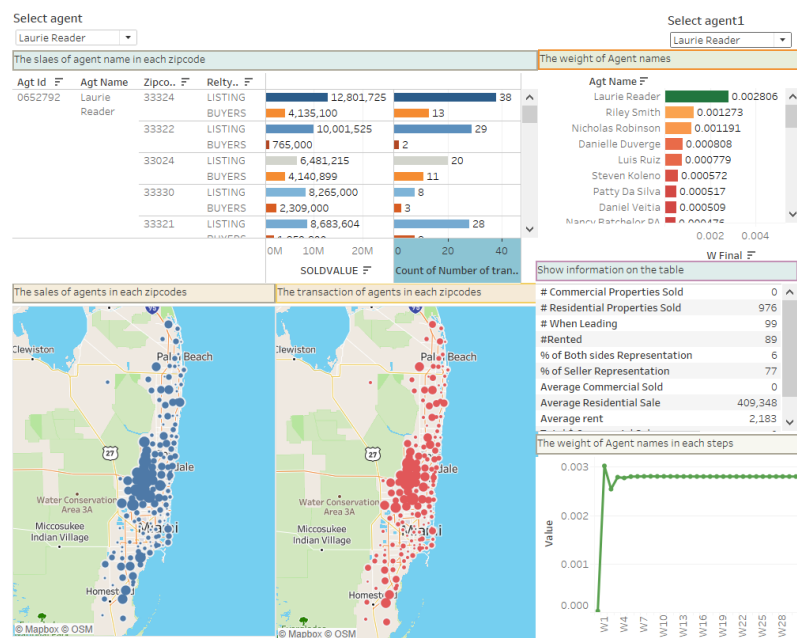


Figure 6. Spend-Based model dashboard



**Figure 7.** Simplified Transaction Based Model

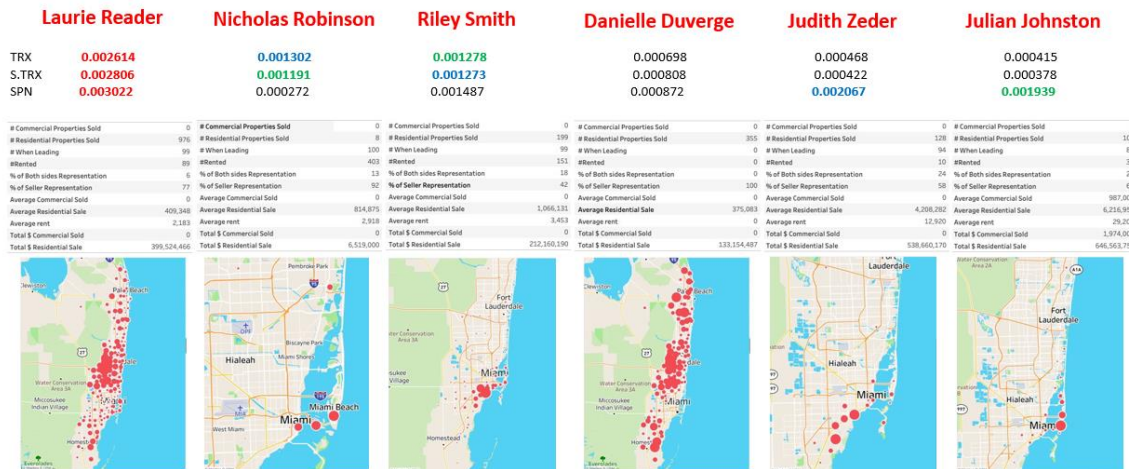
### 3.6 Results

Based on Transaction-Based Model, the top agents are Laurie Reader (0.002614) and Nicholas Robinson (0.001302). Laurie Reader is the most productive agent in South Florida with 976 residential sales and 89 rentals closed in 2021. Nicholas Robinson, the second highly ranked agent, had 403 rentals, but only 8 residential sales closed in 2021.

According to Simplified Transaction-Based model, Laurie Reader is the top-ranked agent, as in the previous model, with score 0.002806, while agent Riley Smith replaced Nicholas Robinson with score 0.001273. It happened because in rental transaction it is more common to have agent representing both sides of the transaction (landlord and tenant) than in sale transactions when agent represents seller and buyer, therefore removal of “loop” links impacted Nicholas Robinson.

When we calculate scores using Sales-Based model, Laurie Reader is still the top-ranked agent with score 0.003002, number two and number three are Judith Zeder and Julian Johnston with scores 0.002067 and 0.001939. Judith Zeder and Julian Johnston are well-known luxury realtors that sell very expensive properties. In fact, the average sizes of their transactions are \$4.2 Million and \$6.2 Million, and Julian Johnston was featured by the Million Dollar Listing show during several seasons.





**Figure 8. Model Comparison – Top Agent**

Each model has advantages and drawback, which make them more applicable to different purposes. We recommend using the Simplified Transaction Model as the easiest for implementation. The selected model is also less influenced by the size of the sales, the transaction frequencies between the same agents, as well as by sales when the same agent represents buyer and seller.

#### 4. Conclusion and Next Steps

The purpose of our study was to apply PageRank algorithm to the problem of ranking the licensed real estate agents. We adapted the original algorithm from Google to the real estate problem and designed three versions of the algorithm to evaluate agent performance based on the strength of their networks. The methodology created new approach to evaluate the agents. It can be a valuable instrument for consumers in selection of which agent to work with.

Each of the three developed models has advantages and drawback depending on the objectives and practical situations. The Simplified Transaction Model is the simplest model, and it is less influenced by expensive properties, and transaction frequencies between the same agents.

In our application we demonstrated that PageRank algorithms can be used outside the original set-up of webpage ranking. The methodology produced meaningful results in the practical application of the ranking of real estate agents. As the next step, we plan to test and validate the results of the agent ranking with buyers and sellers of real estate, as well as consider other practical implementations related to real estate, such us ranking of real estate brokerages and service providers.

## References

- [1] Velma Zahirovich-Herbert, Benie W. Waller, Geoffrey K. Turnbull, “Properties that Sell at or above Listing Price: Strategic Pricing, Better Broker or Just Bumb Luck?”. *Journal of Real Estate Finance Economics*, vol 60, pp. 53-76, 2020.
- [2] Lawrence Page, Sergey Brin, Rajeev Motwani, Terry Winograd, “The PageRank citation ranking: Bringing order to the Web”, Technical Report, 1998
- [3] H. Xu, E. Martin, and A. Mahidadia, “Contents and time sensitive document ranking of scientific literature”, *Journal of Informetrics*, vol. 8, pp. 546-561, 2014.