Project Proposal:

Big Data at AT&T Regional

Fall Case Competition

October 17, 2016

Russell Holloway

Richard Dalrymple

Kenny Ryan

**Table of contents:**

* 1. **INTRODUCTION**
  2. **Background**
  3. **Problem Statement**

**2.0 SECOND-ROUND REQUIREMENTS**

**3.0 THE PLAN**

**3.1 Non-social retail factors**

**3.2 Demographic Profiles & Zip Code Analysis**

**3.3 Tackling Twitter**

**3.4 Google and Yelp Reviews – a Deeper Dive**

**75206 analysis**

**75225 analysis**

**4.0 EXPECTED RESULTS**

**5.0 RECAP**

**1.0 INTRODUCTION**

**1.1 Background**

AT&T is one of the largest integrated wireless, broadband, video, telecommunications and strategic service providers in the world. AT&T is committed to providing a world class, effortless customer experience. As the recipient of J.D. Power’s top overall ranking in the category of Full-Service Wireless Purchase Experience, AT&T prides itself on its ability to recognize and address customer concerns.

In a global marketplace, where constant innovation and customer contact is vital, AT&T must navigate the landscape of traditional call centers, retail store interactions, and now, social media communication. Every day, 500 million tweets are posted in the Twitter universe. These tweets range from interactions between friends to consumer complaints. As a data-driven company, whose mission is to “connect people with their world, everywhere they live, work and play … and do it better than anyone else,” AT&T can utilize information across the many social media platforms to better serve customers.

**1.2 Problem Statement – Part 1**

Now that we have a good understanding around customer sentiment towards store experience in social media, Big Data at AT&T would like to know what specific services AT&T offers that are rated highest by our customers through social media. We would like for the teams to look at data around technician dispatch, product installation, store experience, satisfaction around product usage or any additional factors found through research to get a holistic view of customer sentiment towards AT&T. Teams should also perform location analysis to see if the services and product experience vary by zones or zip codes. Big Data at AT&T wants teams to think outside of the box and get creative with their research. Investigate things like location based demographics to see if that plays a role in AT&T’s overall customer sentiment. Show us differences between social media platforms or AT&T authorized retailers versus corporate retail stores. At the end of the day, AT&T is all about their customers and ensuring our customers are having great experiences. Big Data at AT&T wants to better understand how our community is shaping the perception of AT&T and what recommendations the teams have to improve customer care.

For the final deliverable, Big Data at AT&T would like to see a complete analysis done around the many factors which affect store performance, the data used to perform the analysis, and how the factors differ by locations. If you can, drill down to specific retail stores and list the factors which are hindering their performance. Please include a visualization with your project whether it’s a map or something interactive.

**2.0 SECOND-ROUND REQUIREMENTS**

* Expand search to include data about AT&T’s products and services
* Research key factors outside of social media that can affect customer sentiment and rate them based on influence / effect on customer sentiment
* Create a visual that reinforces your final findings (can be interactive)
* Based off of your findings, recommend next steps on how to improve AT&T’s customer care

**3.0 THE PLAN**

In round one, we developed a plan that gave AT&T a better understanding of what impact social media and online ratings were having on its retail stores’ performances. That plan had three central initiatives: 1.) Develop demographic profiles of social media platform users; 2.) Address how to collect, analyze and act upon data captured from Twitter; 3.) Address how to collect, analyze and act upon data captured from review sites, such as Google and Yelp.

The second round of AT&T’s Case Competition favors a natural development of this prior plan by: 1.) Expanding our Twitter analysis to include additional AT&T services, such as: technician dispatch, product installation, store experience, satisfaction around product usage, and additional factors; 2.) Applying our zip code analysis to showing how AT&T service and product experiences vary by zones or zip codes; 3.) Applying our demographic profiles to show the role of location-based demographics in shaping AT&T’s overall customer sentiment; 3.) Expanding upon our analysis of how AT&T corporate and authorized retail stores differ in customer satisfaction; 4.) Completing our social media platform demographic profiles to show how various mediums differ.

Our goal is to generate an interactive and visual analysis of Dallas’ AT&T retail stores that ranks them based upon the likely impact of social media and reveals factors hindering performance at low-performance stores.

**3.1 Non-social retail factors**

Before delving into our social media analysis, we should preface this by saying we believe attitudes and experiences reflected by social media will only explain a percentage of a store’s performance. There are numerous factors outside of social media that can affect customer sentiment, which we organized into three rated tiers of likely impact.

Tier 1: Store location, what businesses are near store (i.e. grocer, mall), population around store, education of area, product mix offered by store (relative to local demographics), phone plan price options, AT&T phone / cable service quality in area, demographics, income levels of area.

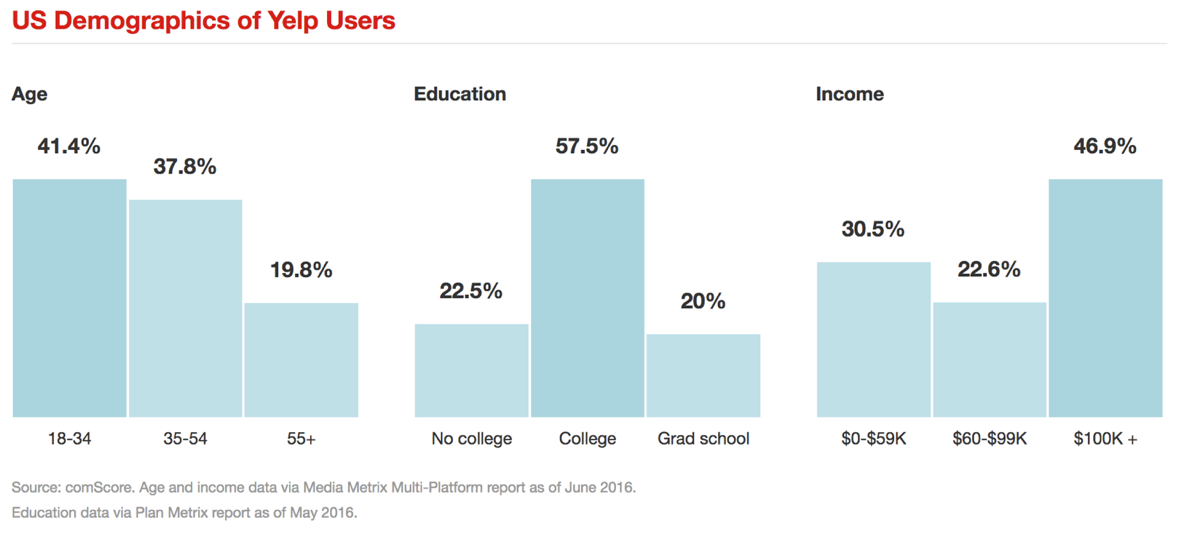
Tier 2: Proximity to highways / major thoroughfares, crime rate, store exterior / curb appeal, traffic patterns, walkability, distance to nearest AT&T store or competitor, availability of parking.

Tier 3: Seasonal sales (spikes when new phones are released or around holidays), weather, politics effect overall retail income: delayed purchases, size of store.

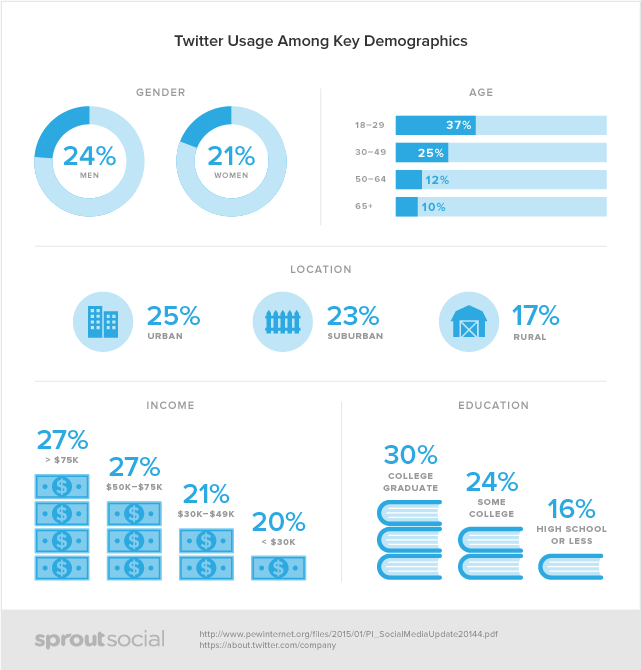
**3.2 Demographic Profiles & Zip Code Analysis**

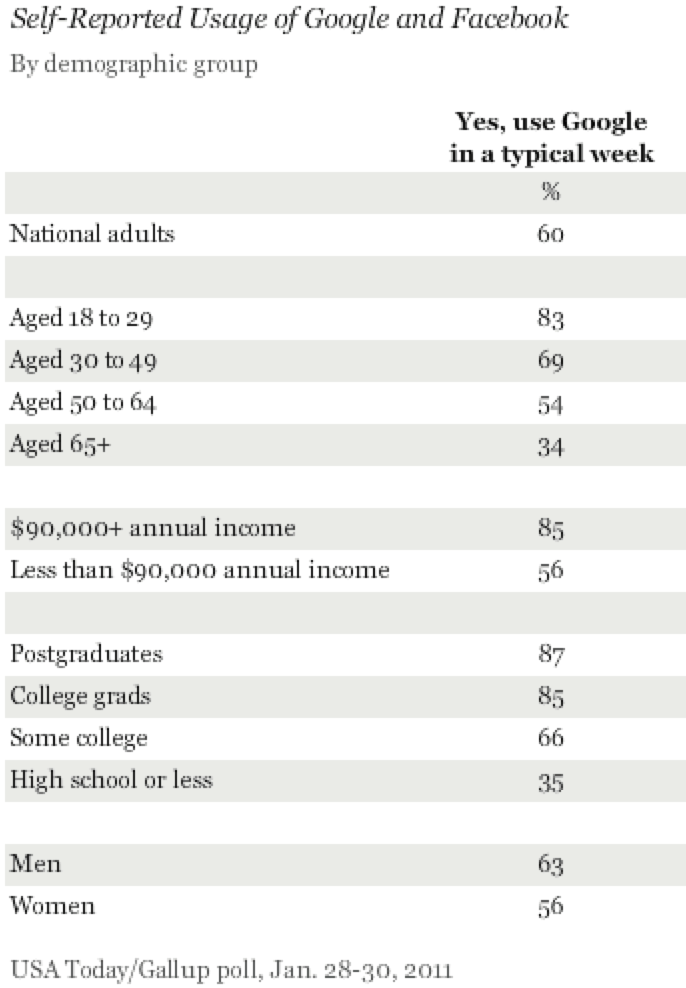
In round one, we showed that Twitter and Yelp users are different. In round two, we’ve created an additional demographic profile for likely Google Review users and codified the differences between all three for convenient analysis.

Data provided by Yelp reveals that its users tend to be college educated, between 18 and 54-years old, and on the higher end of the income spectrum.



Data provided by the Pew Research Center and Twitter shows that Twitter users are young, moderate earners and most haven’t graduated college.



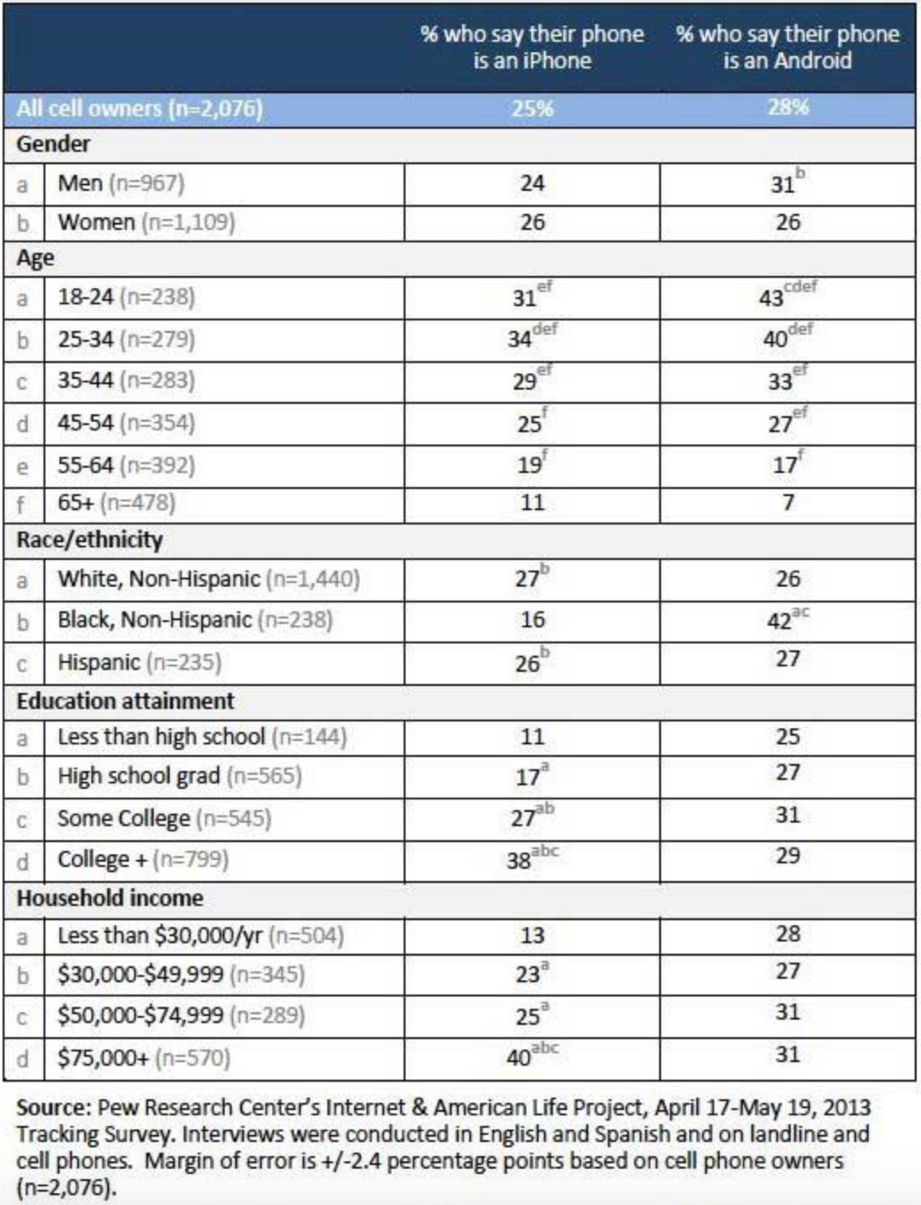
Data provided by a 2011 Gallup Poll of internet users shows that Google users are very similar to Yelp users, but slightly less financially affluent. The distant date of this profile makes it less than ideal, but more-recent data was not available on the short timeline of this project.

Based upon this data, we found the key differentiators to be education and income level.

Yelp and Google users were nearly identically educated, within the margin of error. Thus, our first step is to weight each zip code by the likelihood its residents use Google Reviews and Yelp, based on the zip code’s education level.

According to a quantcast.com report from this summer, Yelp has 96,600,000 unique monthly users across its various platforms. If 77.5 percent of these users have college or graduate degrees (according to previously mentioned yelp statistics), then Yelp has 72,933,000 monthly users with college or graduate degrees. According to U.S. Census data, there were 321,418,820 Americans in 2010, and 29.3 percent (94,175,714) of them had college degrees. Thus, 72,933,000 divided by 94,175,714 equals **77.4 percent of Americans with college degrees using Yelp**. The same methods can be applied to deduce **10.4 percent of Americans who lack college degrees use Yelp** (96,600,000 monthly Yelp users \* 22.5% of Yelp users don’t have degrees / 321,418,820 Americans \* 70.7% of Americans lack college degrees).

The U.S. Census reveals the percent of each zip code’s population that has completed a college degree. Thus, we can weight each zip code to see what percent of its population is likely to use Google or Yelp when doing shopping research (77.4 percent \* percent of residents with college degree + 10.4 percent \* percent of residents without college degree).

The surprising result of Twitter is that, though it skews toward users who are younger, less educated and less wealthy, it is used remarkably consistently between zip codes. Thirteen of the 15 zip codes with an AT&T retail presence have a Twitter penetration rate between 20 and 23 percent. This means that Twitter’s weight can be applied to all zip codes universally, but AT&T should understand that the complaints or compliments of Twitter users likely represent customers lower on the socio-economic order in any given area. It’s also worth noting that Twitter’s American user base is on the decline.

The final differentiator worth mentioning is that, though Google Review and Yelp users are nearly identical, we do believe there could be one key difference: Yelp is the default rating for businesses searched for on Apple maps, which is the default map application of iPhones, and Google Reviews are the default ratings for businesses searched for on Google maps, which is the default map application of Android devices. Apple users do lean toward the wealthier end of the spectrum than Android users, so if AT&T were to wonder which demographic is represented by each social media and rating platform, it would conclude that Yelp is used by mostly upper-class customers, Google Reviews by middle-to-upper class customers, and Twitter by lower-to-middle class customers.

Now we’re going to reveal how data and sentiments are captured and analyzed on the various platforms.

**3.3 Tackling Twitter**

There is one significant challenge to using Twitter as a metric of how social media is impacting AT&T’s performance in any given zip code: Most users block Twitter from sharing their location.

This means that, unless AT&T has an internal method for tracking the points of origin of tweets, most of the commentary about AT&T on Twitter is untraceable, and thus of limited use.

That said, there are techniques that can be used to measure how an area’s residents feel about AT&T.

Twitter content can be searched on a geographic basis. For instance, pick coordinates in the middle of a zip code or city, set an appropriate radius, then search for references of AT&T or AT&T-related terms (such as ATT, #ATT, @ATT, etc.).

Because of the limited number of Twitter users who publish their locations, this is more effective on a larger-than-zip code basis.

Once applied, this will result in a database of area tweets that can be analyzed in two ways.

The first method of analysis is to take the temperature of the area to generate a 1-5 star rating, similar to what you’d find on Yelp or Google Reviews. To generate this temperature, we first looked at 209 Google and Yelp reviews, and generated a list of all 2,953 unique words used in those reviews.

We then organized the unique words by how often they appeared in reviews of each star-value. For example the word “loyal” appeared five times, exclusively in one-star reviews. It’s weighted as a one, because it only appears in one-star reviews. The word “we” appeared in 67 one-star reviews, seven four-star reviews and 21 five-star reviews. It’s weighted as a 3.52, because it appears 67 times in 134 negative reviews, seven times in six four-star reviews and 21 times in 44 five-star reviews.

The values of these 2,953 words are then applied to tweets from the past year to generate an AT&T thermometer rating of 2.3 in Dallas.

As a test to see how accurate this metric is, we applied it to a randomly chosen AT&T store that had 30 written reviews. Our text analysis predicted a 2.43 rating. The actual rating of the store was 2.4.

While this is most useful in creating a city or region-wide Twitter thermometer score, some users do occasional tweet the street name of the store they are praising or deriding in their tweets. Thus, there are some occasions where Twitter provides intel on a single location’s performance.

Twitter can also be used with keyword and ancillary word searches to reveal data specifically about AT&T products and services. Below are a list of keywords (next to black dots) and related ancillary terms (next to white dots) that can be found by applying search functions to a city’s tweets.

* **Technician dispatch**: Technician-installer
  + friendly-rude-slow-late-“where is”
* **Product installation**: technician-installer-installed
  + Slow-wrong-clean-messy-gross-professional-unprofessional
* **Store experience**: service-guy-employee-manager
  + Slow-fast-rude-wait-friendly
* **Product usage**: Internet, GoPhone, iPhone, Nexus, Galaxy, Note7, Note 7, Pixel, cable, blue tooth, head sets, speakers, batteries, case, cases, reception, modem, U-verse,
  + Slow-disappointing-out-broken-down-fast-quick-cheap-expensive-pricey, reception-router

Keep in mind that though we believe this analysis can be very enlightening, it mostly reflects the opinions and experiences of younger and less affluent clients across the city.

**3.4 Google and Yelp Reviews – a Deeper Dive**

The biggest place where location-based demographics are felt on social media is the Yelp and Google Review-sphere. Using our methodology, we have created an interactive, visual ranking of social media’s impact on AT&T’s Dallas stores that can include deep-dive information revealing the “why” behind variances in store performance. This can be used as a source of recommendations on how AT&T can improve customer care. This analysis also reveals how location-based demographics play a role in AT&T’s overall customer sentiment and shows the differences between AT&T retails and corporate stores.

The first step to our methodology is to collect, by zip code, the Google and Yelp review scores of every AT&T corporate store or authorized retailer, which we did with a bit of Python code developed by one of our teammates. This will yield the absolute scores of each store.

The next step of our analysis is, we believe, a critically important one. In order to rank the stores, we take the additional step of measuring AT&T’s performance in each zip code relative to its competition.

The reason for this is that AT&T does not exist in a vacuum. Its ability to attract new sales is dependent upon how it’s regarded in relation to its local competition. We believe the absolute-value list will likely better correlate to cancelations in a zip code, but the relative-value rankings will likely better correlate to sales. For instance, any zip code where AT&T rates lower than two stars might have rapidly dropping retention rates due to customers’ poor experiences when they go to the store for new phones, upgrades, or other services. The relative-value list is likely more important in attracting new customers, as it more accurately measures what a customer sees when they are trying to decide which service provider they want to buy a phone or equipment from in the area.

Thus, the absolute-score list can help AT&T identify stores that must be improved to better retain area customers, while the relative-score list will help AT&T identify which zip codes must be improved to better attract new customers. The relative-score list will likely better correlate to prospective-customer foot traffic, and thus new sales, than the absolute score list, as well.

We believe this is important because there may be some zip codes where AT&T stores average a three-star score, but their local competitors average a two-star score, thus AT&T is the best option in that zip code and can expect a larger percent of foot traffic from customers looking for a new phone company or new phone equipment. On the flip side, there may be other zip codes where AT&T stores average a four-star score, but competitors average a five-star score, so even though AT&T is rated higher in this second zip code than it is in the first on an absolute level, AT&T is relatively the worst option available and thus could expect fewer walk-ins from potential new customers. Thus, the four-star score can rank below the three-star store in expected sales due to differences in area competition.

The methodology for building this relative ranking is as follows: Gather the ratings of each individual phone store in a particular zip code. This list would include AT&T, Sprint, Verizon, T-Mobile, and others. Create one list, “Column A,” that simply lists the given scores of each store, then create a second list that measures the stores on a relative level (the equation would be *(Store X – MIN(Column A)) / (MAX(Column A)-MIN(Column A)) \* 4 + 1*). This second list will grant each zip code’s top-rated store a five and its worst-rated store a one.

Next, Google Review and Yelp scores should be weighted based on our projected penetration of Yelp and Google Review users in each zip code. This is done using the aforementioned method of evaluating zip code income levels (outlined in 3.2). Once weighted Google and Yelp Review absolute and relative scores have been created, they can be summed to create a ranking of Dallas stores.

The next step of our Google and Yelp analysis makes this ranking an actionable source of retail improvement ideas. When Google and Yelp scores are gathered, each location in a zip code should have its reviews analyzed by a program that searches for key words we identified in our Twitter analysis as the most-likely to be commented-upon aspects of the retail-shopper’s experience. Many of these terms exist within synonym groups. For example, when reviewers are commenting upon a store’s customer service, they frequently use terms like “Service,” “Employees,” “Salespeople,” and “Guy,” etc. Once a search has isolated reviews about service, these service-related reviews can be sorted by score (one-star, two-stars, etc.), then a secondary search can be made for ancillary terms. For example, when reviewers comment on a store’s customer service, they often write it’s “Slow,” “Friendly,” or “Rude,” etc.

By applying our ranking and deep-dive analysis methodology to a city or region, AT&T can quickly identify which stores are underperforming in the eyes of social media users, and then diagnose what can be done to turn those locations around.

**4.0 RESULTS**

There are 18 AT&T retail locations located in 15 Dallas zip codes. We have applied our Google and Yelp Review analysis to all of them, and created a city-wide Twitter thermometer score showing the impact of that social media as well. All of this is available in our Tableau package, but for the purposes of this write up, we will provide results when applied to two of Dallas’ zip codes below: 75206 – which contains the city’s highest-ranked store, and 75225, which contains the city’s lowest-ranked store (ignoring stores penalized for not having Yelp reviews).

**75206 analysis:**

The 75206 zip code is home to five phone stores. Two are AT&T authorized retailers, two are Sprint stores and one is a Metro PCS location. The AT&T authorized retailer at 5618 East Mockingbird Lane is the highest-rated AT&T store in Dallas. It possesses two, five-star Yelp reviews and seven Google Reviews for an average Google score of 3.8 – relatively the area’s best. The zip code also boasts a relatively high 58.2 percent of residents with college degrees, meaning it has a population that’s likely to look for those Google Review and Yelp ratings and heed them.

Other information we can tell you about this zip code is that 15.9 percent of its residents are between the age of 20 and 29 – the demographic most likely to be on Twitter – meaning its Twitter penetration is likely an average 21 percent.

The zip code’s population is 36,248. Its income is slightly lower than most of the other zip codes AT&T has a presence in, meaning we’d project 29.9 percent of its population likely prefers android devices and 24.9 percent of its population likely prefers iPhones – with the rest preferring alternative mobile devices.

Because there are relatively few reviews, a deeper dive is optional but not necessary – the reviews can simply be read on their own merit. This quickly reveals that the store has a very strong sales force that is popular with shoppers – though at least one reviewer lamented that, being an authorized retailer, it could not provide all the same deals available in a corporate store.

The takeaway is that employees who are kind and helpful – likely the result of strong hiring and training practices – are one way to improve retailers’ ratings online.

**75225 analysis:**

The 75225 zip code is home to four phone stores. There is one AT&T corporate store, one AT&T authorized retailer, a T-Mobile store and a Verizon store. The AT&T authorized retailer at 5960 W. Northwest Highway is the lowest-rated AT&T store in Dallas. It possesses a single, one-star Yelp review and three decent Google Reviews – but three reviews are not enough for Google to display an aggregate score, so it appears unscored which is equivalent to a zero. Eighty-one percent of the zip code’s residents possess college degrees – more than any other zip code AT&T has a presence in. This means that a relatively large number of residents are likely using Google or Yelp to guide their purchasing decisions, and they are being deterred by the store’s low Yelp score and lack of a visible Google rating.

Other information we can tell you about this zip code is that only 5.2 percent of its residents are between the age of 20 and 29 – the demographic most likely to be on Twitter. Of all the zip codes AT&T has a presence in, this is the one least-likely to be impacted by Twitter.

The zip code’s population is 20,892. Its income is significantly higher than the other zip codes AT&T has a presence in, meaning we project that 34.7 percent of its population likely prefers iPhone devices and 30.4 percent of its population likely prefers android devices – with the rest preferring alternative mobile devices.

Because there are relatively few reviews, a deeper dive is optional but not necessary – the reviews can simply be read on their own merit. The Yelp review is actually a broader complaint about AT&T’s call center customer service, so while it tells us a bit about what area consumers think about AT&T, it’s not particularly insightful about this particular store. The Google Reviews paint a mixed picture, with one reviewer from a year ago complimenting the customer service, but another reviewer from seven months ago complaining about their long wait time in the store.

What are the takeaways for this zip code? Namely, it needs more reviews! If employees at the store were encouraged to ask their friends or family to review their shopping experiences at the store, even a single five-star Yelp review and a single five-star Google review would be enough to make the Google Review score suddenly visible and lift it in our rankings from worst to second-best in Dallas in likely impact of rating sites on sales.

**5.0 RECAP**

We believe this combination of Twitter and Google / Yelp review analysis can help AT&T rank its stores and find the degree to which social media correlates to store performance. To recap how we’ve answered this case study’s challenges:

* **Expand search to include data about AT&T’s products and services***Our Twitter analysis now includes keywords and ancillary terms to collect data on a broad range of AT&T products and services.*
* **Research key factors outside of social media that can affect customer sentiment and rate them based on influence / effect on customer sentiment***We provided a tiered ranking of factors beyond social media that likely impact customer sentiment and retail performance.*
* **Create a visual that reinforces your final findings (can be interactive)***We created an interactive and visual Tableau map that allows AT&T to quickly evaluate all of its Dallas retail locations.*
* **Based off of your findings, recommend next steps on how to improve AT&T’s customer care***Individual stores can be improved based on data revealed by our Yelp and Google Review analytic methods. City-wide performance can be improved by applying a deeper dive to our Twitter analytics.*