

Final Assignment

Deciding a platform to develop web radio software. Does the music type influence the number of visitors attending?

Experimental Methods and Statistics

HCI 5400

Mahmut Erdemli

Introduction

Brief Overview

Web streaming is an increasingly popular method of broadcasting radio on the Internet. The main difference between Internet-based broadcasting and traditional radio broadcasting, besides the obvious difference in terms of accessibility and technology, is that it is the type or genre of content that is transmitted. Hence, the target audience between the two mediums is very different, making it a consumer targeting issue along with a design interface issue. Indeed, consider the fact that global Internet usage as of 2013 is 77% of the population within developed countries (“Global Internet usage,” 2014) with 61% and 75% of total population having access in the Americas and Europe respectively, companies intending to target this young and global audience need to put more focus on designing appropriate scheduling schemes to attract as many listeners as possible.

Nublu was originally a club that played live music between 10pm and 2 am throughout the week. It had a global presence with considerable popularity in wide range of locations such as New York, İstanbul and São Paulo. Recently, the company decided to stream their content through the Internet and created Nublu.fm that is designed to stream on three different servers at three different time settings based on their geolocation. They play the schedules based on their local time or closest distance. When we build this radio, original idea was to let access to a broader audience and promote live music service as bonus value. During the rest of the day, Nublu.fm plays licenced tracks of artists who perform live music at the bar.

The purpose of the current paper research paper is to statistically explore several important aspects of Nublu.fm's scheduling. Specifically, I will explore the following questions: (i) Do participants listen a session more than other ones?; (ii) Is there a significant difference on the sessions during the day then weekends, depending on the music that is playing (session)?; (iii) Is there an effect between individual sessions? (iv) Is there an interaction between live music and traffic hours (morning traffic, rush hour)?; and (v) Where do we receive the most n the number of visitors?

A totally grass roots, independent and vibrant community spirit is the driving idea of Nublu FM. With ongoing process of content creation from the community of Nublu world, and the industry growing attention to online radios, our goal is to expand our listenership with unique programming depending on our demographics.

While the use of web based internet radio is increasing, the demographic of users can let us know who can access the product, where and when. An experimental research on analytics can shape the elements that structure the radio, depending on the use case and the persona. When the content is the focus, in our case it is what kind of genre of music is presented to the user, automatically investment becomes the location-based advertisements. This research is useful to understand how scheduling affecting on the interest of users, how to increase the popularity of the streaming internet radio. The strategy of this research is to measure which program of the radio to invest, and if it is a good idea to let live music stream to increase the number of audience.

Visitors' data is conducted at the original server computer that is also streaming content, checking every individual connection by looking at the hardware settings, this is how we could separate individual and concurrent users from each connection.

Summary of Data File

Day Session Data File:

Session: Individual sessions include live and recorded music,

Dayofweek: Seven days of week,

Visitor: Number of visitors who are connected,

Minutespervisitor: Average minutes per visitor connected,

Weekend: Workdays and weekends(Saturday,Sunday).

Device Session Data File:

Session: Individual sessions include live and recorded music,

Agent: Unique device that visitors use to connect the radio,

ConcurrentConnections: Listeners that are connected in the same time,

Visitors: Number of visitors who are connected,

Averagebandwidth: Average download size in kilobytes,

Seconds: Seconds of music that is being listened,

Minutes: Average minutes per unique user,

Songs: Average number of songs per unique user.

Traffic Hours vs Live Music:

Session: Individual sessions include live and recorded music,

Agent: Unique device that visitors use to connect the radio,

Visitors: Number of visitors who are connected,

Songs: Average number of songs per unique user,

Seconds: Seconds of music that is being listened,

Minutes: Average minutes per unique user.

Methods

Participants

Nublu.fm has been running for the past 867 days with a total of 111,507 songs that have been playing. Nublu.fm has been visited 11,345 times by 3,412 individual visitors in which 2819 of them used their personal devices, either through mobile and desktop.

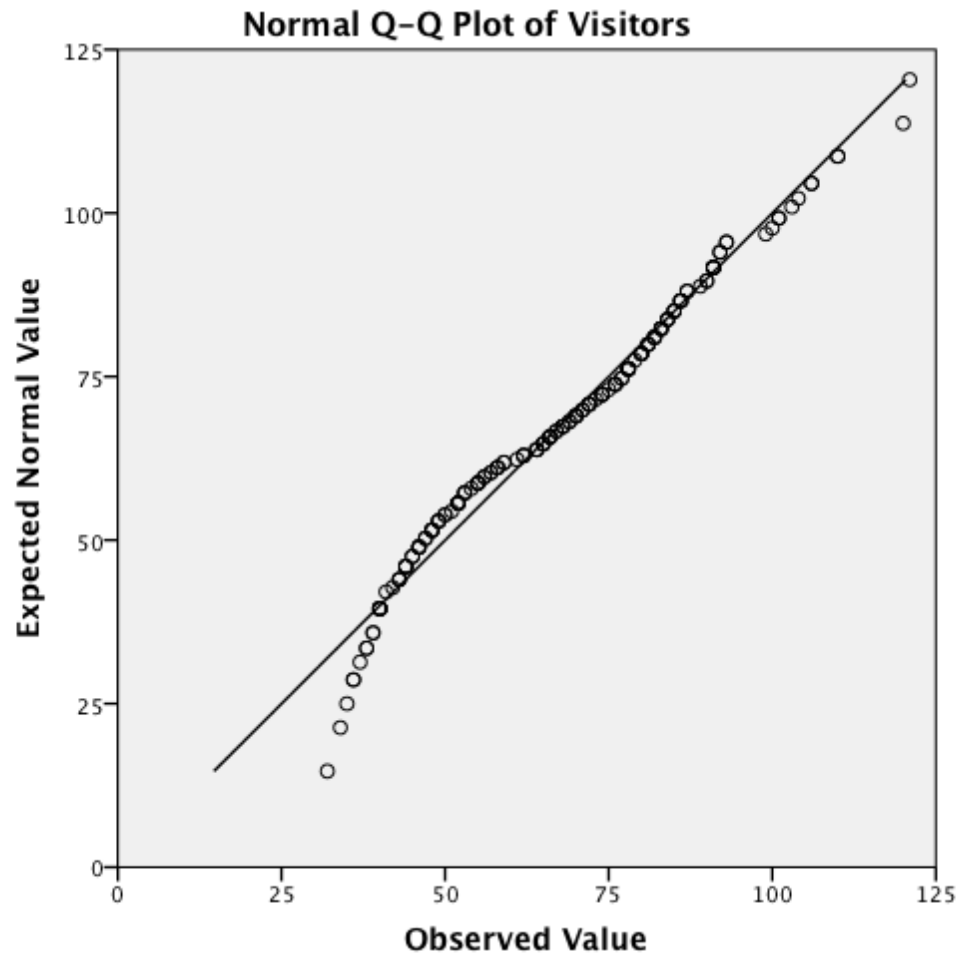
Procedure

There are three data files with *day session data file* observed in a chronological order. Audience who are connected to individual sessions on daily basis are counted through a digital counter. Furthermore, Nublu.fm has eight different sessions that build the schedule. (i) live music from 10pm to 2am, (ii) night critters from 2am to 6am, (iii) morning traffic from 6am to 9am, (iv) Office hour from 9am to 12pm, (v) lunch break from 12pm to 2pm, (vi) wasting time @ work from 2pm to 5pm, (vii) rush hour from 5pm to 7pm and (viii) prime time from 7 pm to 10 pm. *Device session data file* observes each individual personal device users connection, on two different groups, mobile and desktop. *Traffic vs live music data file* observes on listeners connects on a chronological order.

Results

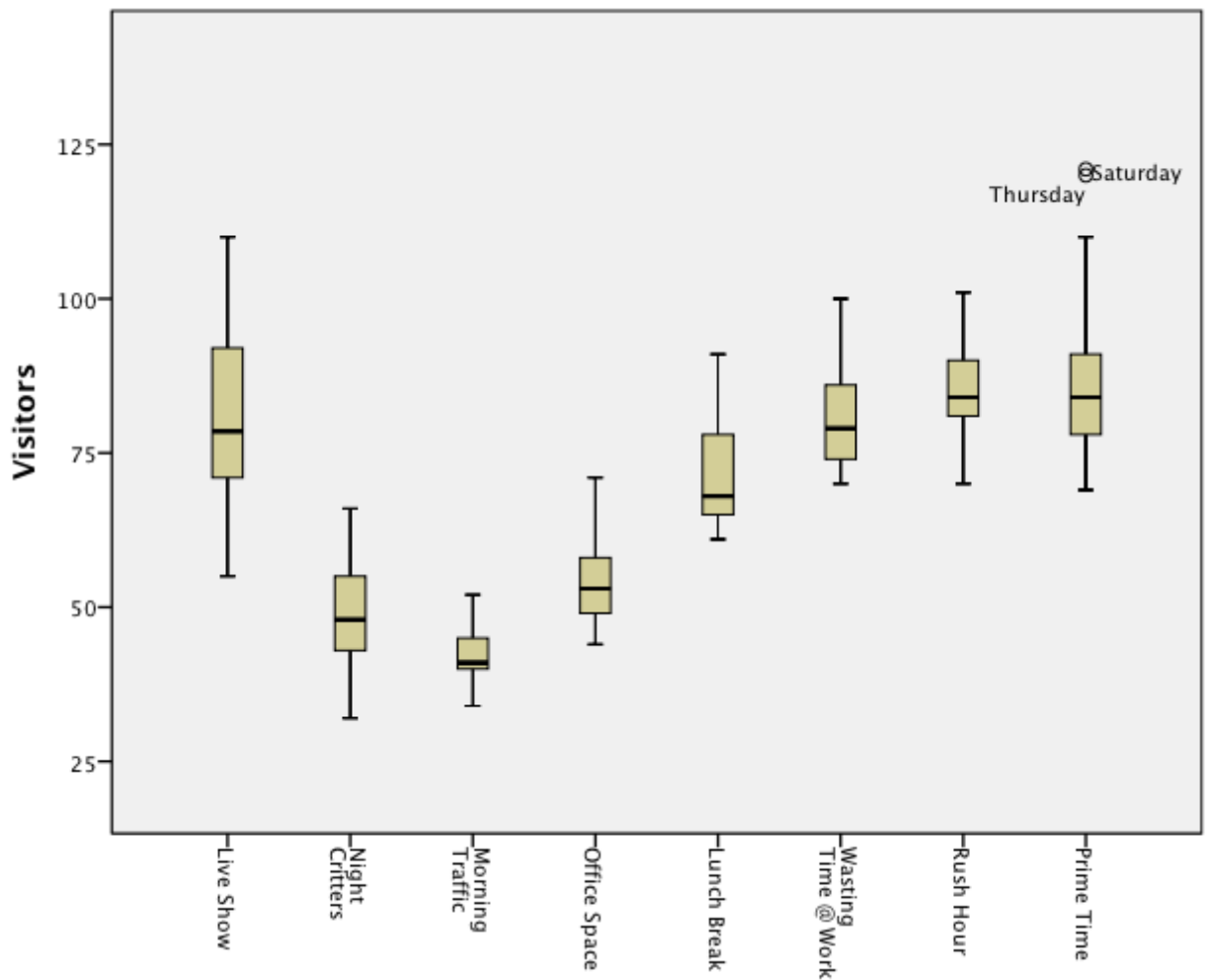
To begin with general descriptive, the mean value for visitors is 67.53 ($SD = 19.744$) ($n=168$), Kurtosis $-.589$ $SE = .373$, Skewness 2.55 $SE = .187$. Mean for minutes per visitor is 47.6583 ($SD = 14.496$), Skewness 3.114 $SE = .187$, Kurtosis 11.684 $SE = .373$.

Figure 1. Q-Q Plot for average visitors in general



When we look at the visual representation of the visitors data on a scattergram we can see the correlation relations helps identify problems such as a curved relationship or the presence of outliers.

Figure 2. Histogram divided by sessions that repeat each day



There are minimum number of outliers which means that outliers do not distort the inference that I obtain from the correlation coefficient. There are only two data points on a Thursday and Saturday that are relatively outlying but, given the sample size, should not have a significant impact on the statistical inference.

Figure 3. Mean average of visitors in total during the day

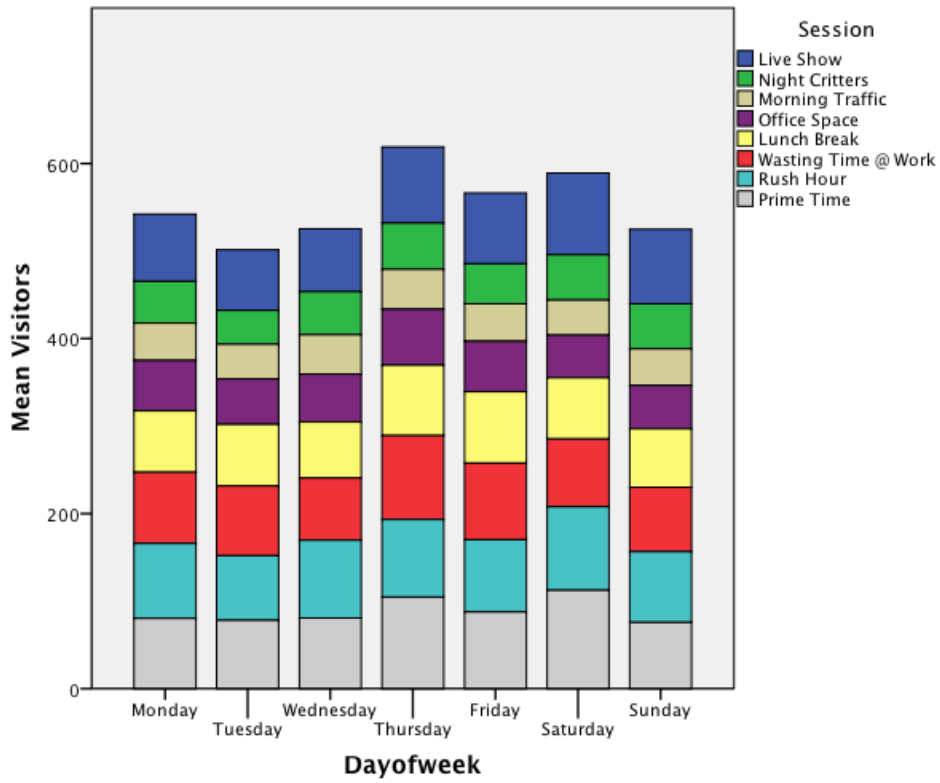


Figure 4. Mean average of visitors per session during the day

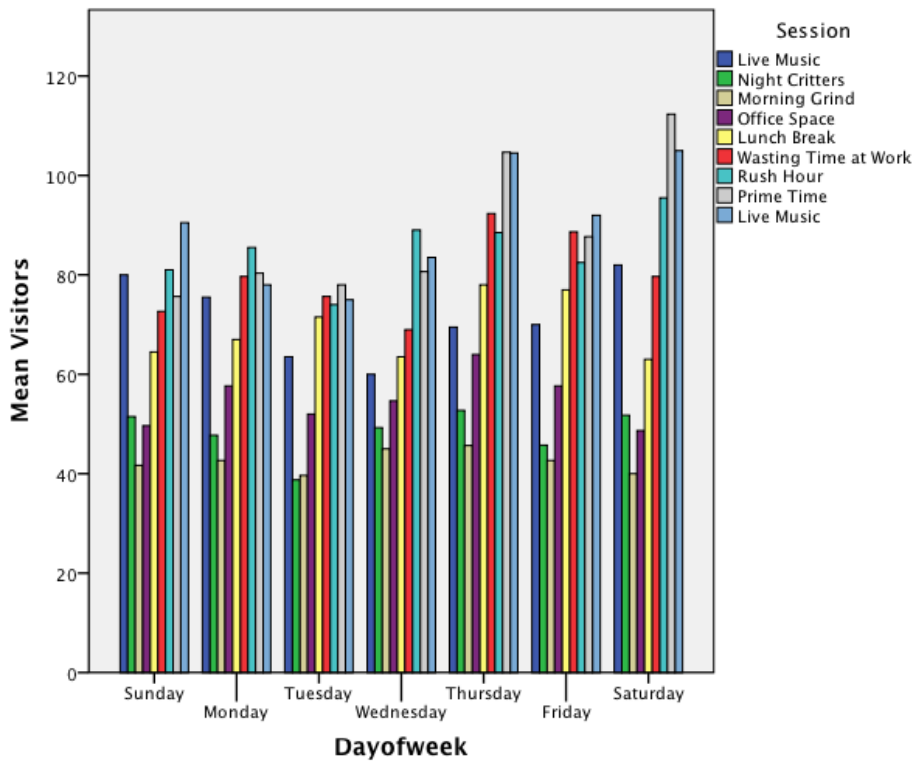
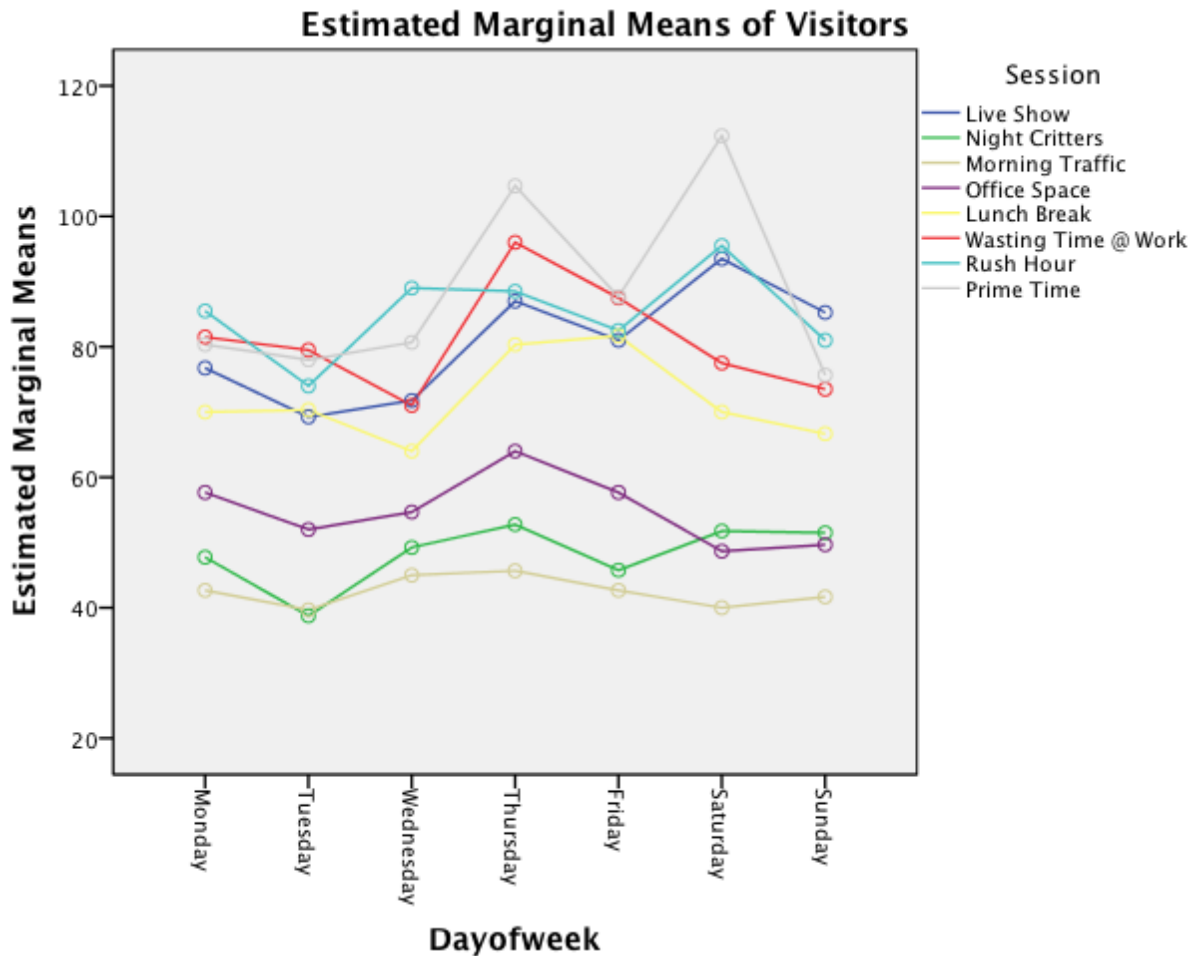


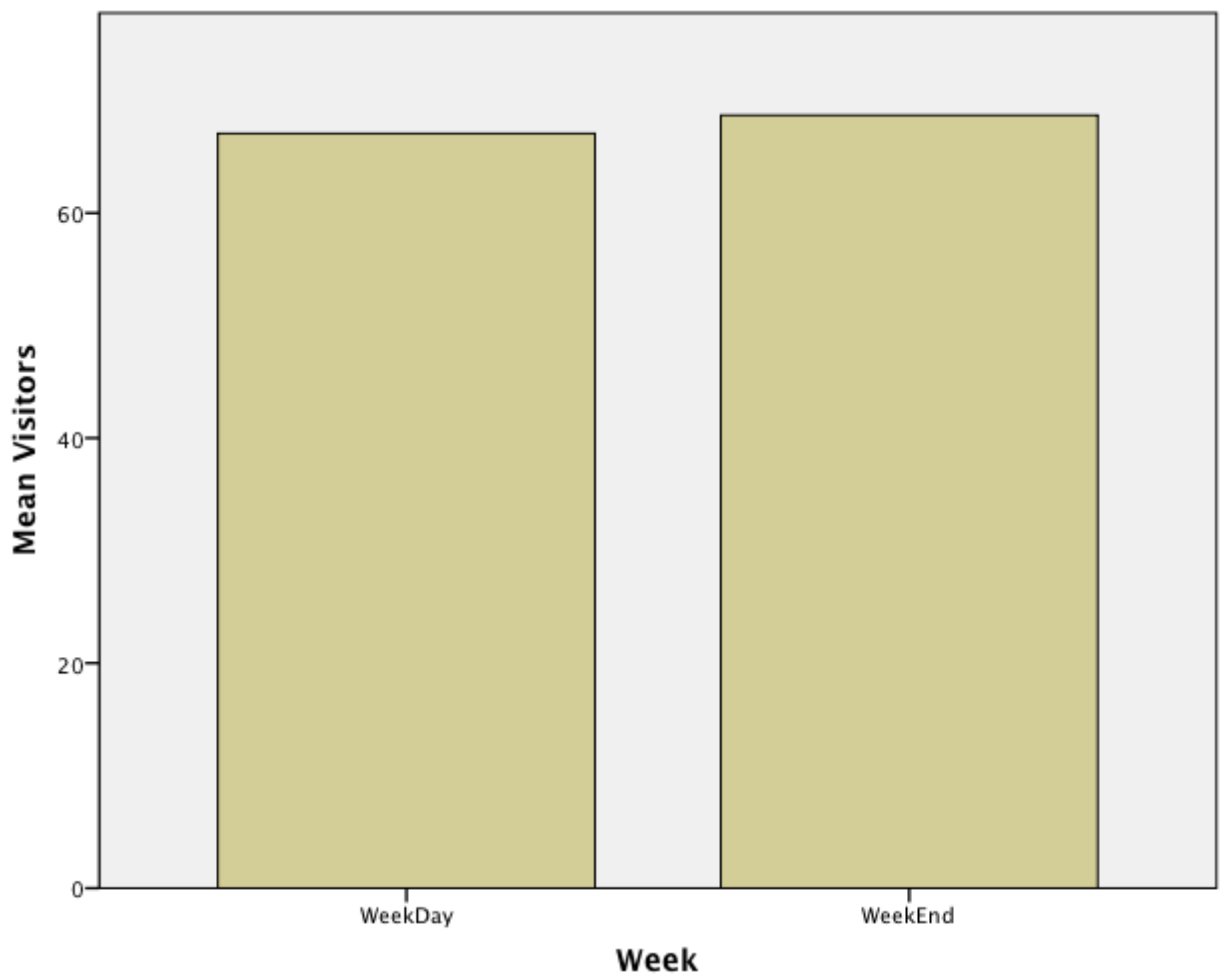
Figure 5. Line graph indicating average visitors per session on each day



The average mean for visitors connected to the radio on Monday ($M = 66$ $SD = 16.650$), is higher than Sunday, the average visitors to listen to the radio ($M = 64.88$ $SD = 16.645$), $t(46) = .234$, $\eta^2 = 0.069002835$, two-tailed $p = .816$. Tuesday is the least visited day ($M = 60.79$ $SD = 17.080$), followed by Wednesday ($M = 64.04$ $SD = 15.584$), $t(46) = -.683$, $\eta^2 = -0.201405712$, two-tailed $p = .498$. Thursday recorded the second highest number of visitors connected to the radio ($M = 75.50$ $SD = 22.485$). The t-test for average number of visitors for Fridays ($M = 69.00$ $SD = 19.715$) and Saturday ($M = 72.50$ $SD = 25.972$) shows that Saturday has the highest average number of visitors $t(46) = -.526$, $\eta^2 = 0.217329443$, two tailed $p = .602$.

Next, a two-way unrelated ANOVA shows that there are significant effects obtained for our pre-scheduled sessions, $F(7, 112) = 97.391, p < 0.001$, partial $\eta_p^2 = .859$, and days of week, $F(6, 112) = 8.445, p < .001$, partial $\eta_p^2 = .311$, and a significant interaction on the day of week between the sessions on a day, $F(42, 112) = 1.729, p = .012$, partial $\eta_p^2 = .393$. The highest number of listeners of live music is on Saturdays, ($M = 93.50 SD = 13.404$), and the most listened times are the prime time session Saturdays ($M = 104.67 SD = 14.572$), then on Thursdays $M = 112.33 SD = 7.767$. The visitors who listen the rush hour session go up to $M = 95.50 SD = 7.778$ on Thursdays and morning traffic session can go low as $M = 39.67 SD = 5.132$.

Figure 6. Average number of visitors on weekend and weekdays



To compare the means of visitors that connect to the radio on a basis of weekdays and weekends, I ran an independent sample t-test on SPSS. According to Levene's test $F(1, 166) = 1.874$, $p = .173$ is not statistically significant which shows for equality of variance assumption holds. The mean number of connected visitors on weekdays ($M = 67.07$ $SD = 18.881$) is significantly higher, $t(166) = -.480$ $\eta^2 = -0.07451045$, two-tailed $p = .632$, than the listeners who connect on the weekends ($M = 68.69$ $SD = 21.921$).

Next, I looked into differences in terms of the mean number of minutes connected to Nublu.fm between weekend and weekday basis to see whether there is a significant difference. Because the variances for the two groups were significantly different ($F = 47.536$, $p < .001$), a t-test of unequal variances was used ($t(50.136) = 1.525$, two tailed $p > .005$). The results demonstrate that the mean of duration of listening time per visitor on weekends ($M = 51.5694$ $SD = 24.48261$) is significantly higher ($t(20) = -2.81$, two-tailed $p = .011$), than minutes per visitors during the week ($M = 46.0938$ $SD = 7.03695$). Hence, we can confidently conclude that, on average, listeners tend to connect to Nublu.fm for longer periods of time per session during the weekend than on the weekday.

The subsequent test performed a two-way unrelated ANOVA to show that there was no significant effect obtained for the difference of variance between the number of visitors on work days and weekend ($F(1, 152) = .198$, $p > 0.005$), with a small effect size partial $\eta_p^2 = .001$. But there was a significant effect obtained for number of visitor difference in sessions, $F(7, 152) = 61.544$, $p < .001$, and a big effect size, partial $\eta_p^2 = .73$. I also found a significant presence of interaction, $F(7, 152) = 2.570$, $p < 0.5$, and a big effect size partial $\eta_p^2 = .106$. However Levene's test shows the variance of errors are not homogeneous by having a significant value $F(15, 152) =$

3.876 , $p < 0.1$. Hence, while I do not find any significant difference between weekend and workday in terms of number of visitors, I do find significant differences in number of visitors across sessions.

Relatedly, I looked at one-way unrelated analysis of variance and obtained an overall significant effect for the visitors connecting to different sessions ($F(7, 160) = 65.476, p < 0.001$). Scheffé's test found that the number of visitors for Live Show session is significantly different from that of the morning traffic session, $p < .001$, the office space session, $p < .001$, and the night critters $p < .001$. Another group that shows significant difference is morning traffic session. It is differed with multiple sessions live show, lunch break, rush hour, prime time with $p < 0.001$, and office hour by $p < 0.5$.

Extending the previous test, I ran live music session as the control group and found a significant two tailed result at visitors who listen live music, this control group it shows a significant difference between the other sessions ($t(160) = -6.193, p < 0.001$). And for contrast group two, with coefficients assigned as -7, 1, 1, 1, 1, 1, 1, 1, we can say live music control group has a significant negative value of effect on the other groups. Rush hour and morning traffic sessions are thought of the time of the day we are driving, and if we set our as our control groups with a coefficient setting of 1, 1, -3, 1, 1, 1, -3, 1, we get s significant value of the two groups showing an effect of a positive value($t(160) = 3.528, p = .001$). Finally if we run a contrast group of solely the morning traffic group, with a coefficient assigned as 1, 1, -7, 1, 1, 1, 1, 1 there is a positive significant effect of impact on the other groups ($t(160) = 12.646, p < 0.001$).

When we run a two way ANOVA, first we need to understand whether the variance within the three groups is significant. I will choose the groups that show the

biggest effect sizes which are morning traffic, rush hour and live music. We need to see whether the Levene's Test of Equality of Error variances is not significant. We need to see if the error variances are significantly different and if the distributions are the same or they are equal between the groups, In this case the data is not homogenous, with significant value $p < .001$. Therefore we need to run post-hoc test Tukey.

When we run a post-hoc test, according to Student-Newman-Keuls test, office space and lunch break are the only significantly different sessions from the other preplanned sessions. $p < 0.001$. When the data is drilled down to a visitor based on individual device that is being used with the mean frequency of visits per individual user is 2.74, ($SD = 3.659$) ($n = 2819$). And a positive Kurtosis, 41.566 $SE = 0.92$ value, Skewness, 5.410 $SE = .046$. SPSS compare distribution of visitors with unique device, since it is a positive value we can see the distribution is pretty steep with a high value.

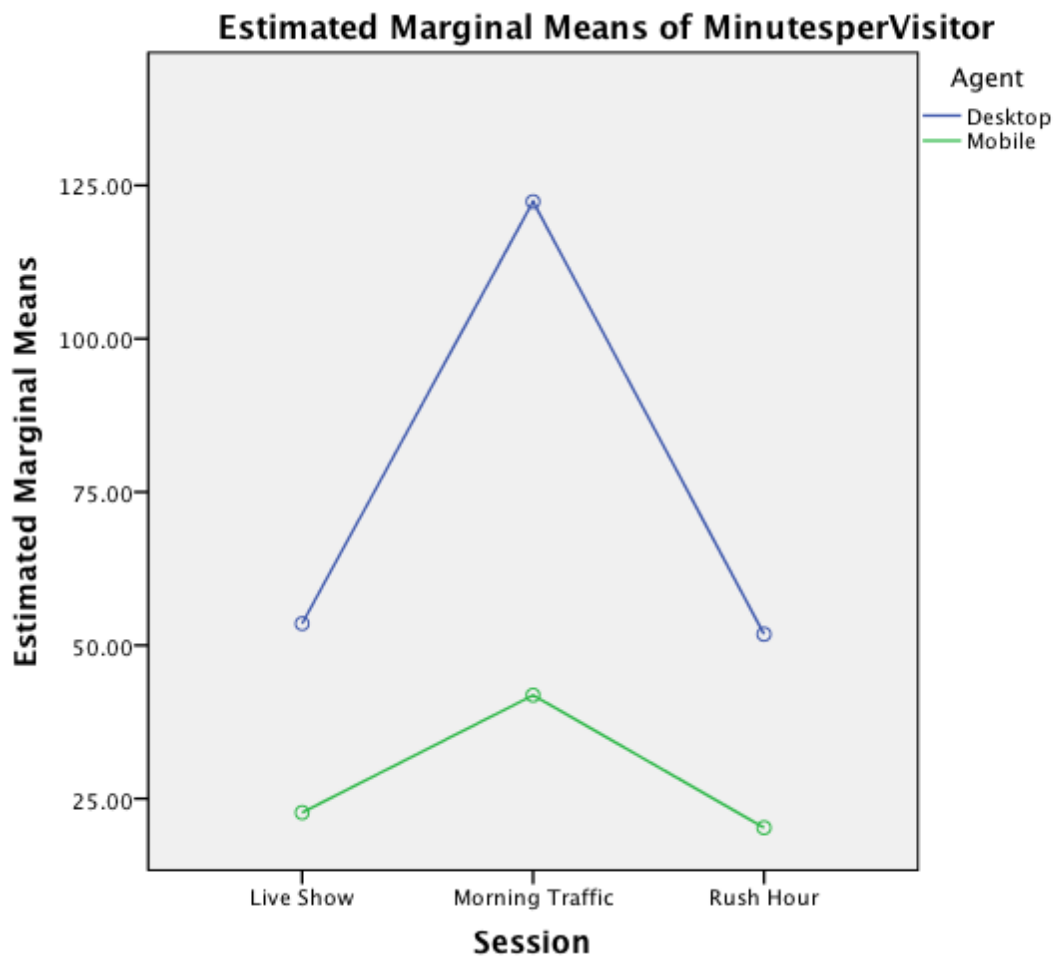
The duration that the radio is being listened per each visitor by the mean is 56.26 minutes visitors per hour ($SD = 157.458$) ($n = 2819$). SPSS calculated Kurtosis to, see the steepness of the and as minutes per listener the distribution curve is pretty steep, 418.312 $SE = 0.92$, Skewness, 19.013 $SE = .046$. The songs has been listened per visitors is 10.836, $SD = 10.1425$ songs per visitors ($n = 2819$) with the same steepness of distribution in Kurtosis, 418.312 $SE = 0.92$, Skewness, 19.013 $SE = .046$ since minutes are calculated of the duration of the songs.

Concurrent connections are the individual listeners that are connected to the server at the same time. This is helpful to decide if there are necessary updates to be done on the hardware, as increasing the number of servers. The mean average for the concurrent users on Nublu.fm at an hourly basis is, 4.82 ($SD = 3.370$) $SE = 0.46$,

Kurtosis 5.257 $SE = 0.92$, Skewness 1.859 that tells us the concurrent listeners connected with their personal device are variant.

When we look at the duration of visitors listening the radio from their desktop, ($M=2926.93$ $SD=10810.802$) versus mobile ($M=1279.94$ $SD=2450.699$) the difference is 1646.99 seconds. Because the variances for the two groups were significantly unequal, $F = 14.009$, $p < .001$, a t-test $t(2877.725)=6.581$, two tailed $p < .05$ for unequal variances was used. The 95 per cent confidence interval for this difference is 2137.709 to 1156.266. Since this interval does not include 0.00, the difference is statistically significant at the two-tailed .05 level.

Figure 7. Profile Plot between the traffic sessions and agent



Since we are planning to invest on high quality streaming of live music, comparing live music sessions with two groups that are chronologically overlaying the data we might obtain during traffic hours, which are rush hour and morning traffic sessions can help us observe the data to see if the radio is being listened on a desktop system or via their mobiles at traffic and live music hours. Identifying an individual device helps to define individual users. To get the best results, I optimized the data to another separate SPSS data file, traffic hours vs live music, and examined a two-way unrelated ANOVA and there was a significant effect for selected sessions, $F(2, 1053) = .889, p < 0.03$, partial $\eta_p^2 = .011$. However there is no effect of unique personal device that a visitor uses to get connected, that are called agents, $F(2, 1053) = .019, p > .05$, partial $\eta_p^2 = .011$. And there was no significant effect of interaction found between the traffic hours against, live music and prime time, $F(2, 1053) = 1.220, p > .05$, partial $\eta_p^2 = .002$. When I looked at the chi-square value between agents and sessions, there was no significant association found between different individual devices connected and attendance for different types of radio session, $\chi^2(2) = 13.976, p = .052$. Finally, I ran a one-way ANOVA with associated contrast coefficients as -2 for live music against, 1, 1 morning hours and traffic hours combined. Again, the result is a non significant value of the two groups with an effect of a negative value ($t(1056) = -1.160, p = > .001$).

Discussion

There is a significant interaction on the day of week between the sessions on a day. Saturday nights are the most popular time that users listen to our sessions. Monday mornings are the least popular days that listeners are listening to our playlists. Furthermore, there is a significant interaction between the workdays and the weekends. There is strong evidence that listeners are interested in live music, especially during the weekend. We can also say that the mobile users have a strong interest for the radio during the traffic hours, which confirms that optimizing the quality for the mobile platform can be beneficial to increase the amount of listeners. While this experimental research is strong in differentiating live music fans, the limitations of this study are that the radio is active but not advertised. With the market out there, we are on hold to process and promote the Nublu FM brand, and we wish to communicate better towards the general public, as well growing our audience. Therefore for researches in this area it is advised to focus on the increasing number of visitors that will add everyday. It is necessary to further research on the issue of promoting the radio while this research is valuable in terms of choosing the content.

References

- Global Internet usage. (2014, December 20). In *Wikipedia, the free encyclopedia*. Retrieved from http://en.wikipedia.org/w/index.php?title=Global_Internet_usage&oldid=635929744
- Ignatov, D. I., Nikolenko, S., Abaev, T., & Konstantinova, N. (2014). Online Recommender System for Radio Station Hosting: Experimental Results Revisited. In *2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)* (Vol. 1, pp. 229–236). doi:10.1109/WI-IAT.2014.38
- Keith, A. (1932). How radio influences human attitudes. *Religious Education*, 27, 198. Retrieved from <http://search.proquest.com/docview/1296941572?accountid=9894>
- Radio Revs Unchanged Year-To-Year: \$17.6 Billion For 2013. (n.d.). Retrieved December 21, 2014, from <http://www.mediapost.com/publications/article/221475/radio-revs-unchanged-year-to-year-176-billion-f.html>
- Verkasalo, H. (2010). Mobile Audience Measurements in User Experience Research. In *2010 IEEE Wireless Communications and Networking Conference (WCNC)* (pp. 1–6). doi:10.1109/WCNC.2010.5506573

Appendix A

Recent market research for digital broadcasting based on U.S.

The US radio advertising revenue is estimated to be around \$17.6 billion for 2013, and digital broadcast within that revenue has been the year over year winner in terms of growth. Although digital accounts for less than 5% of total revenue at \$889 million, it has grown 16% from 2012 to 2013, with an expected rise in 2014 and further ahead. (“Radio Revs Unchanged Year-To-Year,” n.d.)

| Revenue Comparisons - 2013 vs. 2012 (In Millions) | | | | |
|--|----------|-------|----------|-------|
| Revenue | \$Q4 '13 | % Chg | \$FY '13 | % Chg |
| Spot | 3,626 | -3% | 14,054 | -1% |
| Network | 294 | +7% | 1,122 | -4% |
| Digital | 244 | +18% | 889 | +16% |
| Off-Air | 438 | +11% | 1,584 | +5% |
| Grand Total | 4,602 | Flat | 17,649 | Flat |

Source: Miller Kaplan Arase LLP*

Source: Miller Kaplan Arase LLP

According to Mary Meeker (former Morgan Stanley analyst and current Kleiner Perkins investor) streaming music content in the US has risen to 118B units consumed, a growth of 32% over 2013. In Europe, the growth of digital may not be as strong as the US, but according the Radio Advertising Bureau of UK, 55% of Londoners now listen to digital radio weekly (that's 6.5 million listeners), with digital estimated to be growing at a 3.7% rate year over year since 2011.

Reliable figures for Asia Pacific are perhaps even harder to obtain, but it is clear that the transition to digital is more or less in motion in this region too.