Trend in US gyms businesses

The Mavericks

(Kushang Bhatt, Sujay Sangalikar, Darshil Varia, Purvish Oza)

**Introduction**

Fitness is strong focus of today’s lifestyle; the fitness company promises to deliver the highest level of personal service and attention to its member. The objective of this project to define a trend in gym business to improve the current and future gym business and which facilities businesses must provide for getting higher rate.

**Problem Statement**

Which days and what time is customer prefer to go to gym? Which facilities help businesses to get higher rating and does location matter to ratings?

**Tools and Technology**

Java - jdk 1.8.0\_25

Python 3.5

PyCharm Professional 2017.1.1

Winutils

Spark 2.1.0

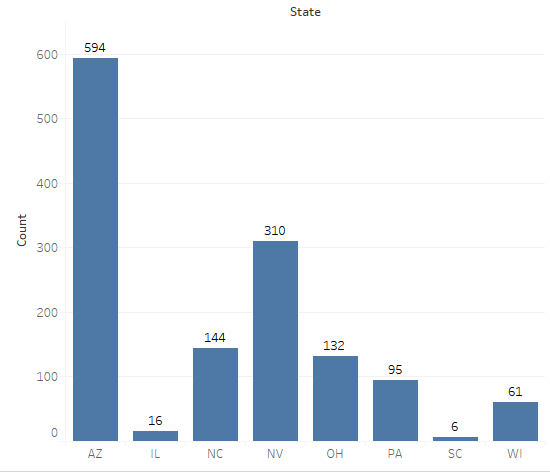
Tableau 10.2.1

We have used PySpark for data extraction and data analysis. We have used Spark in Local mode. Tableau is used for data visualization.

**Methods**

**Data Extraction**

Yelp dataset contain data of 61K businesses for U.K., Germany, Canada and US country. We need only gym related data for US country only. We extracted gyms related businesses from business file for US country only and use those business ids to fetch data from other files. We store this gym related data in a separate file and we have used those file our analysis.



**Figure 1. Total number of gyms per state**

**Data Analysis**

**Which Day and what time customer prefer to go to Gym.**

We use the checkin and business file for this analysis. First, we get the time from list of times when business get maximum number of check-ins. After getting that we divide the DataFrame time column into three column Day, Hour, Check-In and join this DataFrame to business DataFrame to gate the state. We create two list one contain weekdays while other weekends to analyses trend in weekdays and weekends. We create three list for hours.

Weekdays=[**'Mon'**,**'Tue'**,**'Wed'**,**'Thu'**,**'Fri'**]  
weekends=[**'Sat'**,**'Sun'**]  
H1=[**'0'**,**'1'**,**'2'**,**'3'**,**'4'**,**'5'**,**'6'**,**'7'**,**'8'**,**'9'**]  
H2=[**'10'**,**'11'**,**'12'**,**'13'**,**'14'**,**'15'**,**'16'**,**'17'**,**'18'**]  
H3 = [**'19'**,**'20'**,**'21'**,**'22'**,**'23'**]

We started analysis by filtering data by weekdays and hours. First which weekdays and at what hour from H1 different state get maximum check-in.

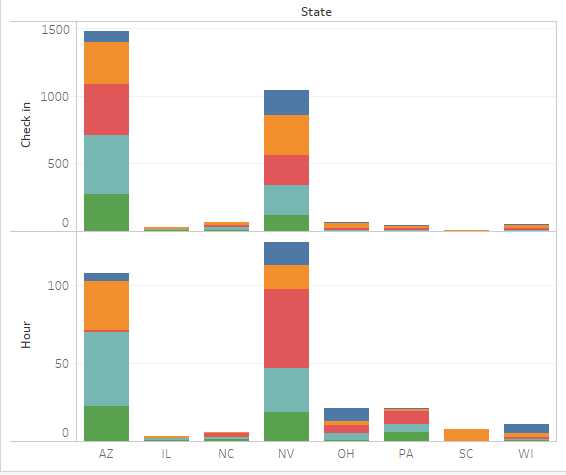


Figure 2 shows that people in AZ state prefer Tuesday, Wednesday, Thursday and Friday and very few people prefer Monday. People in AZ state prefer early morning hours. While people in NV state prefer Monday, Tuesday, Wednesday and Thursday but very few people prefer Friday. People in NV state prefer morning 5 - 8 hour.

**Figure 2. Maximum number of check-in by weekdays and H1 for all state**

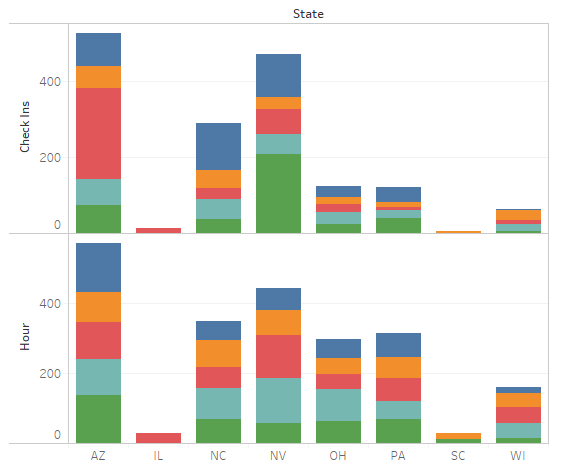


Figure 3 shows that people in AZ state prefer Tuesday, Monday, Thursday and Friday and many people prefer Wednesday. People in AZ state prefer evening 1 - 6 hours. While people in NV state prefer Monday, Tuesday, Wednesday and Thursday and many people prefer Friday. People in NV state prefer evening 3 - 6 hour.

**Figure 3. Maximum number of check-in by weekdays and H2 for all state**

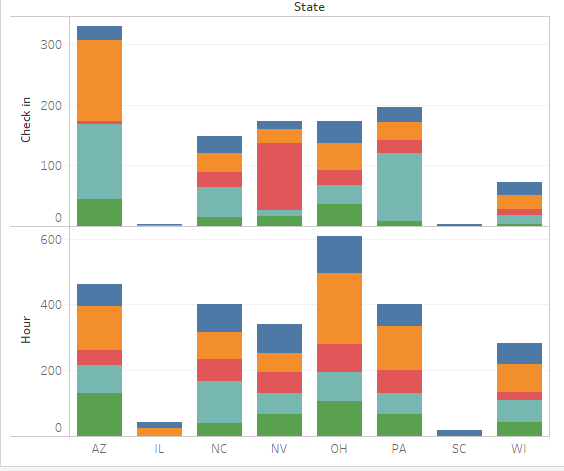
****

Figure 4 shows that people in AZ state prefer Tuesday and Friday and few people prefer Monday, Wednesday and Friday. People in AZ state prefer late evening 17 – 18 and 23 - 24 hours. While people in NV state prefer Wednesday and few people prefer Monday, Tuesday, Thursday and Friday. People in NV state prefer evening 20 - 23 hour.

**Figure 4. Maximum number of check-in by weekdays and H3 for all state**

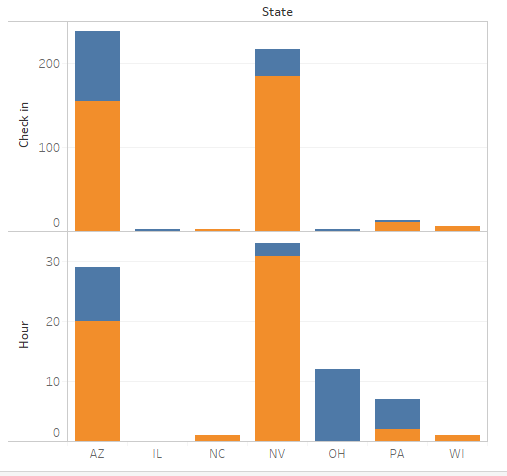


Figure 5 shows that Saturday is the day when people prefer to go to gym. AZ people prefer morning 0 – 4 while NV people prefer morning 1 – 6.

**Figure 5. Maximum number of check-in by weekends and H1 for all state**

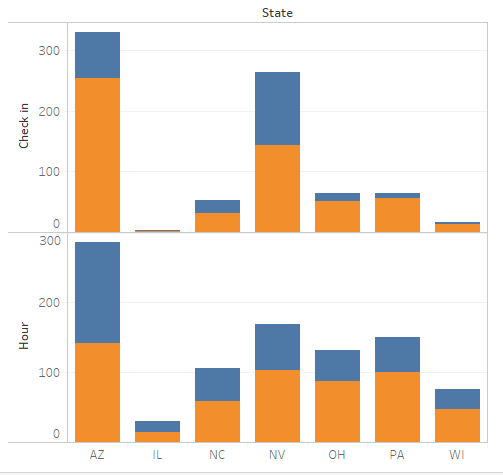


Figure 6 shows that Saturday is the day when people prefer to go to gym. AZ people prefer afternoon 2– 5 while NV people prefer afternoon 2 – 7.

**Figure 6. Maximum number of check-in by weekends and H2 for all state**

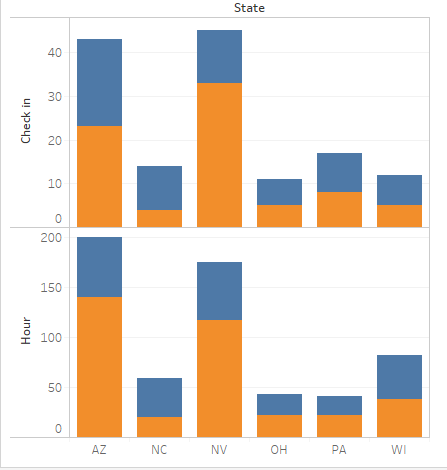
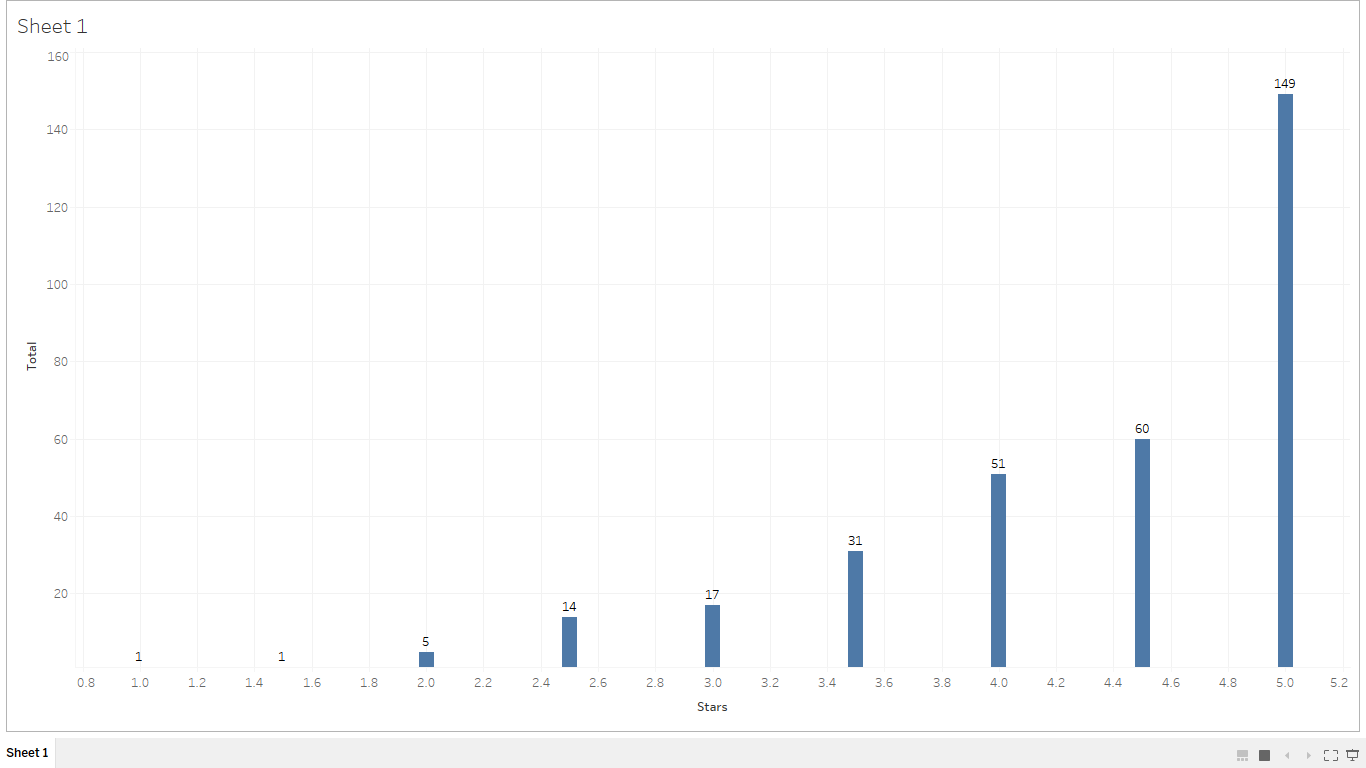
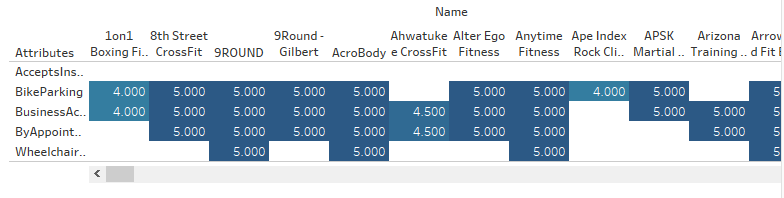
****

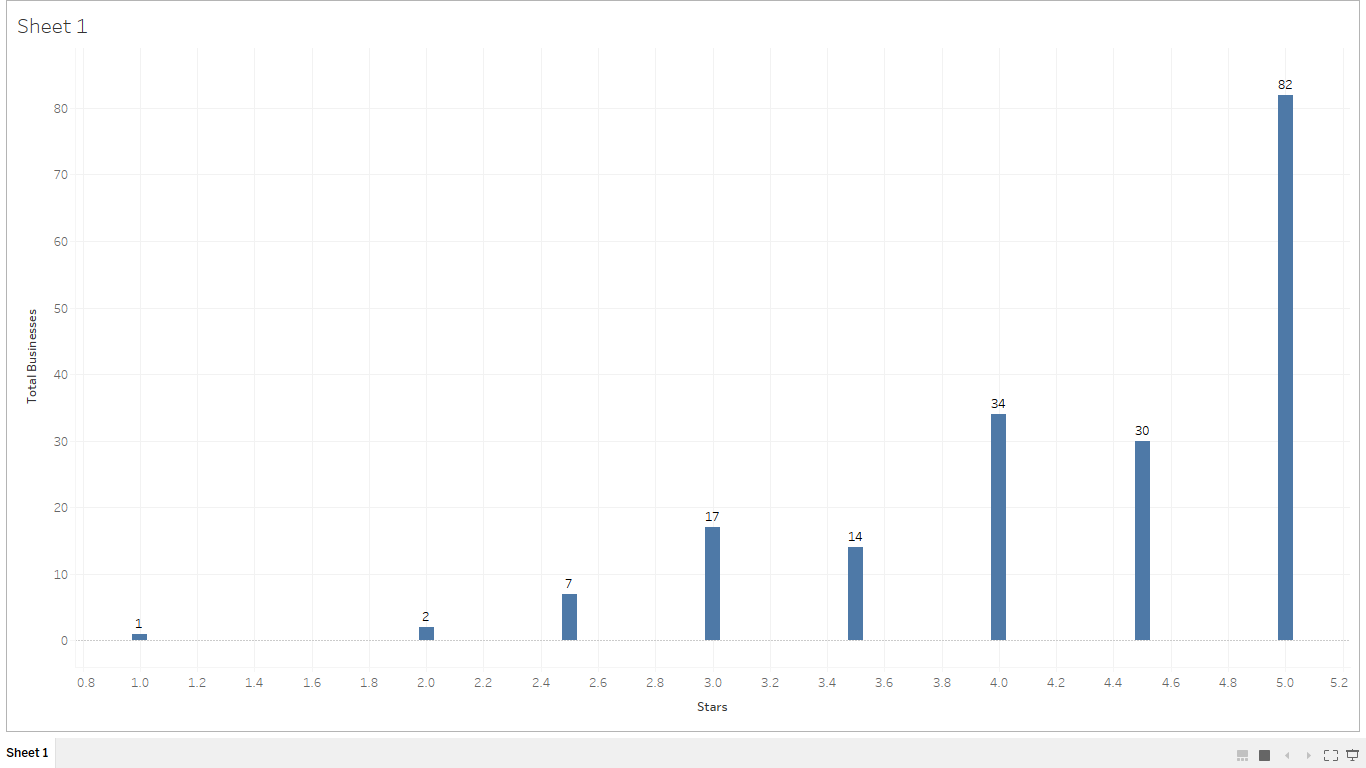
Figure 7 shows that Saturday is the day when people prefer to go to gym. AZ and NV people prefer evening 7– 9.

**Figure 7. Maximum number of check-in by weekends and H3 for all state**

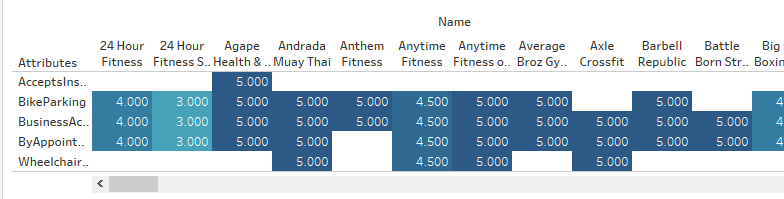
**How attributes impact on business rating?**

**Figure 8: It shows maximum businesses in AZ getting higher rating based on attribute**

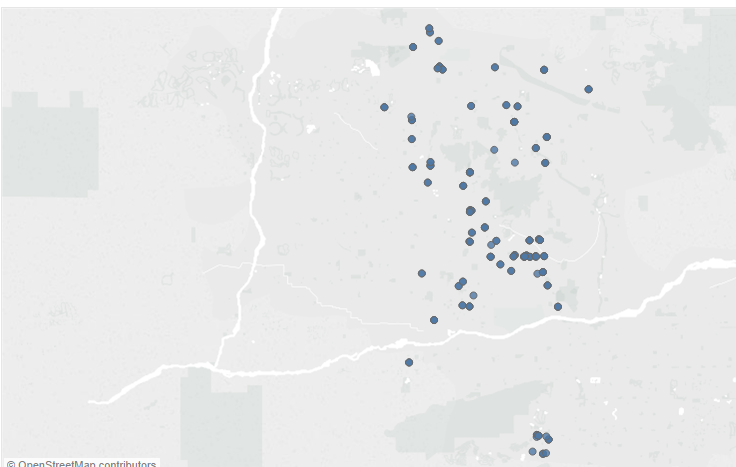


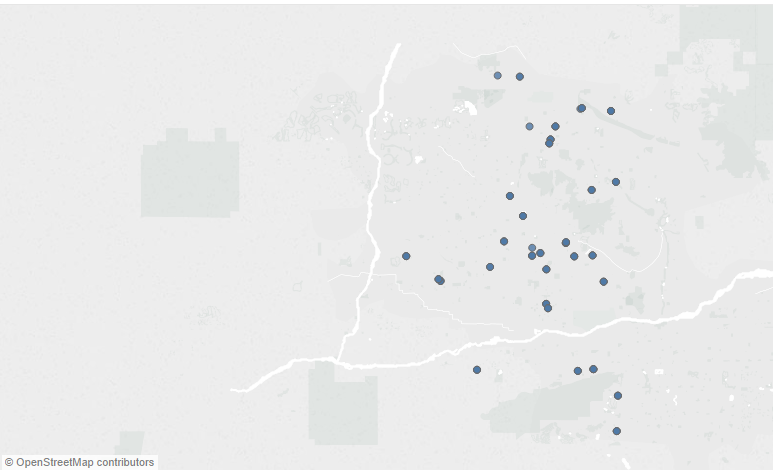


**Figure 9 shows maximum businesses in NV getting higher rating based on attribute**



Businesses having rate greater than 3.5 in AZ Phoenix city are clustered while below 4 rating businesses are scattered. So, location also make an impact on business rating.

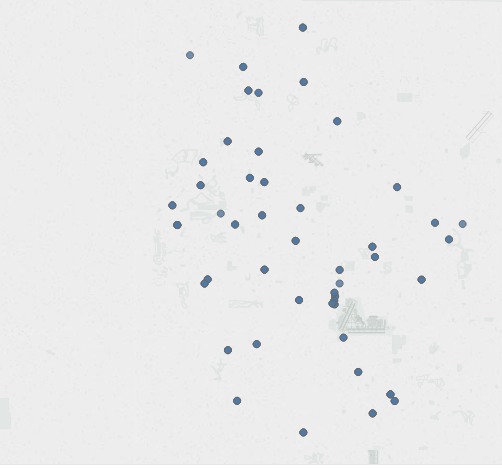
**Figure10: Phoenix top rated businesses are clustered. **

**Figure11: Phoenix least rated businesses are scattered.**

Businesses having rate greater than 3.5 in NV Las Vegas city are clustered while below 4 rating businesses are scattered. So, location also make an impact on business rating.

****

**Figure 12: Las Vegas top rated businesses are clustered.**

****

**Figure 13: Las Vegas least rated businesses are scattered.**

**Result**

* People prefer early morning and late evening in weekdays for gym.
* People prefer evening and mostly Saturday in weekends for gym.
* Bike parking, Business accept credit cards, without appointment, accept insurance and wheelchair accessibility is top facilities customer prefer.
* Location also make impact on business ratings.
* This result help current and future businesses to improve and expand their business.

**Resources**

* <https://mapr.com/blog/easy-journey-from-pig-to-spark-for-developers/>
* <https://cambridgespark.com/content/tutorials/interactively-analyse-100GB-of-JSON-data-with-Spark/index.html>
* <http://spark.apache.org/docs/2.1.0/api/python/pyspark.sql.html>
* <https://databricks.com/blog/2015/02/02/an-introduction-to-json-support-in-spark-sql.html>