**Team: Dataflow**

**Project 1: Soccer Stats**

For this Project, we were interested in finding the correlation between soccer matches and data that might reflect the behavior of soccer fans. As passionate followers of soccer, we Mexicans might be behaving differently during these days and thus this might be reflected in some variables. We decided to explore the relation between soccer matches and four variables for which found meaningful data: prices of beer, DUI detentions, ER emergencies and deaths attributed to homicides and accidents.

The following are the main findings of the project:

1. The prices of beer seem to react to soccer matches. Our approach to measure this was twofold:
   1. During the 2010 and 2014 World Cups, we compared the beer inflation distributions. We bootstrapped beer price inflations during weeks when Mexico played in the World Cup and regular weeks and compared both distributions. **The results show that beer inflations during World Cups are in general higher than in regular weeks, and also that the distribution during WC weeks has fatter tails.**
   2. We compare inflation of beer prices during Clásico Regio weeks, from 2010 to 2015. We make the comparison between prices in Monterrey and the rest of the country. **Results show that average inflations are lower during Clásico Regio weeks than in the rest of the country. This is irrespective of brand, packaging and store chain.**
2. There is a slight relation between soccer matches and DUI detentions. We cross DUI detentions in Monterrey with days where Clásico Regio and Mexico National Team games are held.
   1. **Results show that during Clásico Regio days, DUI detentions do not seem to increase considerably.**
   2. **However, DUI detentions slightly increase the Mexico National team wins.**
3. With respect to ER emergencies, we cross the total of emergencies in Nuevo León in months where Clásico Regio is played twice, due to playoffs or finals of Liga MX. The idea is that such games attract most fans in a city where soccer is fiercely followed. **The main finding is that ER emergencies tend to moderate on match days, but tend to increase the day *after*.**
4. Deaths by homicide and accidents show different behavior when soccer games are held. The approach for both cases was bootstrapping the total amounts of both type of deaths during the second half of 2016 and the first half of 2017, for days when a Liga MX is held and days in which there are no games. **Results show that deaths by homicide seem to slightly moderate during match games, but the distribution seems similar to regular days. On the other hand, deaths by accidents seem to increase during match days, and also the distribution looks different than in regular days. This evidence looks stronger than homicides.**

During this Project we learned important lessons. First, that data collection and cleaning must not be underestimated, as it can be very time consuming and requires attention to detail. Second, in order to fully answer the questions we make, we need to be careful in the analysis chosen and justify it. Finally, it was our first deep experience dealing with all the tools we have learned so far. It has been a great experience and it pushed us to work hard and learn more about the tools we used.