

2017 NBA Hackathon – Business Analytics Track Prompts

Please choose one of the following two prompts to tackle for the next 24 hours. One prompt is representative of league-level business problems; the other of team-level business problems. There is no judging preference for either prompt. Feel free to read through each prompt – and scan the data associated with each – to evaluate which is the best fit for you and your team.

Each team has been assigned a formal “coach” from an NBA team’s business staff who will be around for the beginning of the competition. Members of the NBA Business Analytics team will also be around for the full duration of the competition to answer technical questions, questions related to the prompts and/or questions about their corresponding datasets.

Happy hacking! We are excited you’re here and look forward to reading your submissions.



Submission Details

- A **one-pager** (PDF or Word) is due at 10:30 p.m. in your team Box folder
 - This should serve as an abstract, detailing your planned project, methods and data usage
- The **final submission** is due at 9:00 a.m. in your team Box folder
 - *The recommended submission window is between 8:00 and 9:00 a.m.*
 - Please submit as a **compressed archive** titled [TEAM NAME] Hackathon Submission.zip including the following files:
 - The **one-pager** originally submitted at 10:30 p.m.
 - An additional **project summary** – no more than a few pages – that outlines your project, what it accomplished and which data you used
 - A PowerPoint or PDF (max 5 slides) which will become the **finalist presentation**, should your team be selected
 - A template is available in the 2017 Hackathon Prompts and Template folder
 - A directory titled `Code` which should contain all your code;
 - A directory titled `Miscellaneous` containing images, GIFs, etc. that are supplementary to your project.
 - Please also load your final submission’s compressed archive to your team’s **USB drive**
 - Members of the NBA Business Analytics team will collect them as backup

Option #1 – League-Level Prompt: Entertainment Value of a Game

Relevant Data:

- See `League_Entertainment_Value.zip` on your team's USB drive
 - Complete dataset descriptions and data dictionaries are available in `League_EV_Data.xlsx`
-

Rewind the clock – it's NBA All-Star Weekend 2017 again. There are 460 games remaining in the regular season, and the NBA wants to think strategically about how the league and its teams can optimally allocate promotional spend. To help with its considerations, the league would like to better understand the entertainment value of each game.

Build a model to predict the “entertainment value” of the remaining games. To do so, you'll first need to define what it means for a game to be “entertaining.” This is intentionally left to your discretion, and is important to clearly define and justify. As a North Star, remember that the output of the model is intended to factor into decisions made by the league and its teams to allocate marketing budget. Consider using some combination of the following objective data types:

- TV ratings – data representative of ratings via national or regional sports networks (RSNs)
- On-court metrics – scores, ties, lead changes, etc.
- Web traffic surrounding game day
- Attendance
- Anything publicly available online – social data (e.g., Reddit), published articles, etc.

The regular season extends through mid-April, and situations may change significantly over time. **Thus, the league challenges you to include a measure of how confident you are in your prediction for a given game.** How you approach and communicate this is also at your discretion.

To help train your model(s), we have included the full NBA schedule and results dating back to the 2010-11 season, plus additional data sources that you may consider as direct input features or you may use to derive your own. Such data includes All-Star selections, top jersey sales by year, and metrics representative of tickets sold by game cut at various “snapshot” dates. Beyond the data we have provided, we encourage you to explore open-source resources online such as:

- Historical weather data
 - [Google Trends](#) data
-

Additional notes:

- A lockout in 2011 caused the 2011-12 regular season to begin on 12/25/11; each team played 66 total regular season games
- The New Jersey Nets relocated to Brooklyn prior to the 2012-13 season, but are denoted as `BRK` throughout the data
- The New Orleans Hornets were renamed the New Orleans Pelicans prior to the 2013-14 season, but are denoted as `NOP` throughout the data
- The Charlotte Bobcats were renamed the Charlotte Hornets prior to the 2014-15 season, but are denoted as `CHA` throughout the data

Option #2 – Team-Level Prompt: Leveraging Customer Behavior

Relevant Data:

- See `Team_Customer_Behavior.zip` on your team's USB drive
 - Complete dataset descriptions and data dictionaries are available in `Team_LCB_Data.xlsx`
-

You're the newest member of Team X's BI team. Team X has been collecting data on its customers and their ticket purchases for several years now, but wants to better understand their customers' behaviors – and better act on their tendencies. Teams of all sports around the world attack this same general problem, and now it's your turn to dig in.

Using the given representative datasets, what can you learn about your current customers and non-buying prospects – and what can you infer about their future behavior? This problem is intended to be open-ended. Below we list several sub-prompts to get you started. You are free and encouraged to attack any or as many of the sub-prompts below as you would like; you're also welcome to come up with brand new ideas and put them into action. You can dig deep into one or a couple ideas, or take a wider approach to get a start on several different ideas. The goal is to put you and the rest of Team X in a great spot to move forward after you recover from this Hackathon and work continues next week.

- 1) Identify customer clusters that could benefit Team X, then assign some or all known customers to your clusters. What actions could Team X take to “personalize” the experience of each cluster to drive revenue?
- 2) Build a model or models to predict:
 - a. probability of attendance of a given ticket buyer at a given game, and/or
 - b. total scans for a given game.Can you further predict attendance rates in specific sections or regions of your arena? How can Team X leverage the output of these models – what actions or new offerings might you recommend to the ticket sales team?
- 3) Analyze historical arrival times via the provided scans data. Are there particular scanners or regions of the arena that are currently overloaded – and if so, over any specific time frames? Further, can you predict arrival times for different games at a micro- or macro- level – including specific seat and/or scanner locations? What would you recommend given your findings? Feel free to use the provided `Team_X_Arena_Map.png` file and corresponding `Scanner_Locations_Map.csv` file to visualize your findings.
- 4) Build a model to predict the ticket purchase behavior of a given customer. Are there particular games and/or ticket products you can recommend for a given customer? The term “ticket products” generally refers to the type of bundle – or lack thereof – of a given sale, such as an individual game, mini-plan (e.g., 4-11 games), half plan (e.g., 21 games) or full season ticket membership. Such products could also include perks like food or merchandise vouchers.

For any of the above ideas – or any of your own – feel free to bring in any publicly available outside data you believe could be relevant.

Good luck!