**DATA SOURCES:**

1. Bid Data (2-10%)
2. No bid logs
3. Kochava
4. Agg Data (location data)
5. Census data
6. 3rd party segments (oracle)
7. 1st party data (audible)

**PROCESSES:**

1. App API look up: App level information
2. Store visitations
3. Lat/long validation clean up
4. Bid and no bid data for 30 days
5. IP to zip code (Max mind)

**SEGMENTS**

1. App List
2. Site Descriptions (keyword search) – expanded app list
3. Language preference
4. Topic models that are relevant
5. Census data
6. Location Visitations for segments

**TOPIC MODELS**

1. LDA
2. Patterns of the apps

**RECOMMENDATION/LOOK LIKES**

1. iOS and Android
2. ALS algorithm in Spark
3. User Matrix and Product Matrix
4. If a user didn’t have an app, would he be liking the app or not? For every device and app, we would predict a rating value

**VALIDATION**

1. Kochava gives 80% or more of their data is all Android
2. Nielsen data is used for validation
3. We don’t get sales data, installation of app data ??

**LIVE RAMP**

1. Get identity link for free