**OVERVIEW OF THE “DATA SCIENTIST, ANALYTICS” ROLE:**

**General job description**

Facebook's mission is to give people the power to build community and bring the world closer together. Through our family of apps and services, we're building a different kind of company that connects billions of people around the world, gives them ways to share what matters most to them, and helps bring people closer together. Whether we're creating new products or helping a small business expand its reach, people at Facebook are builders at heart. Our global teams are constantly iterating, solving problems, and working together to empower people around the world to build community and connect in meaningful ways. Together, we can help people build stronger communities — we're just getting started.

We’re looking for Data Scientists to work on our core and business products (ex. Instagram, Messaging, Growth, Engagement, Ads) to help shape the future of what we build at Facebook. You will enjoy working with one of the richest data sets in the world, cutting edge technology, and the ability to see your insights turned into real products on a regular basis. The perfect candidate will have a background in a quantitative or technical field, will have experience working with large data sets, and will have some experience in data-driven decision making. You are focused on results, a self-starter, and have demonstrated success in using analytics to drive the understanding, growth, and success of a product. This position is based full time in our Menlo Park, CA office.

**Responsibilities:**

* Apply your expertise in quantitative analysis, data mining, and the presentation of data to see beyond the numbers and understand how our users interact with both our consumer and business products
* Partner with Product and Engineering teams to solve problems and identify trends and opportunities
* Inform, influence, support, and execute our product decisions and product launches
* The Data Scientist Analytics role has work across the following four areas:
* Product Operations
  + Forecasting and setting product team goals
  + Designing and evaluating experiments
  + Monitoring key product metrics, understanding root causes of changes in metrics
  + Building and analyzing dashboards and reports
  + Building key data sets to empower operational and exploratory analysis
  + Evaluating and defining metrics
* Exploratory Analysis
  + Proposing what to build in the next roadmap
  + Understanding ecosystems, user behaviors, and long-term trends
  + Identifying new levers to help move key metrics
  + Building models of user behaviors for analysis or to power production systems
* Product Leadership
  + Influencing product teams through presentation of data-based recommendations
  + Communicating state of business, experiment results, etc. to product teams
  + Spreading best practices to analytics and product teams
* Data Infrastructure
  + Working in Hadoop and Hive primarily, sometimes MySQL, Oracle, and Vertica
  + Automating analyses and authoring pipelines via SQL and python based ETL framework

**Minimum Qualifications:**

* 2+ years experience doing quantitative analysis
* BA/BS in Computer Science, Math, Physics, Engineering, Statistics or other technical field
* Experience in SQL or other programming languages
* Development experience in any scripting language (PHP, Python, Perl, etc.)
* Ability to communicate the results of analyses with product and leadership teams to influence the strategy of the product
* Understanding of statistics (e.g., hypothesis testing, regressions)
* Experience manipulating data sets through statistical software (ex. R, SAS) or other methods

**Preferred Qualifications:**

* Advanced degrees
* Experience with distributed computing (Hive/Hadoop)

**What the “Data Scientist, Analytics” role is NOT**

* Working in a centralized service organization that responds to ad-hoc requests from product teams. Data Scientists in this role are embedded in product teams and expected to be involved in driving both the high-level direction and the day to day progress of the team.
* Coding and training machine learning models from scratch. There are opportunities in the organization for machine learning, but these roles typically focus on identifying opportunities for new or improved ML applications, experiment development and evaluation specifically for ML models, and performance evaluation for ML models.
* Focusing primarily on ETL, pipelining, and other data infrastructure tasks. The Data Engineering team is focused specifically on this part of the company's data needs.
* Taking a rigorous academic approach to Facebook data and potentially publishing white papers. The Core Data Science team focuses specifically on this area.

**OVERVIEW OF THE INITIAL INTERVIEW, SAMPLE QUESTIONS & ADDITIONAL RESOURCES FOR PREPARING:**

**The interview will be 2 parts (for a total of 30-45 minutes):**

1 – Analytical (10-20 minutes)

1 – Technical (10-20 minutes)

**Analytical Portion (10-20 minutes)**

Per our conversation, the analytical case study will focus on questions to gauge your **product sense**.

For the analytical questions, the interviewer is trying to understand how you solve business questions and problems, as well as how creative and articulate you are at thinking through these problems while solving them. It’s not about the arriving at the perfect, or correct answer, but how you engage with the problem.

Spend some time engaging with Facebook products less as a user and more as someone who is tasked with improving or developing these products. The link below outlines what we consider a “Facebook product” [Ads, Mobile, Timeline, News Feed, Messaging, etc.].

*Resources:*

[Facebook Products](http://newsroom.fb.com/Products)

[Facebook News](http://newsroom.fb.com/news/)

[VP of Analytics, Alex Schultz's Talk at Stanford on Growth at Facebook](https://www.youtube.com/watch?v=URiIsrdplbo)

[How Facebook used Science and Empathy to Reach two Billion users](http://bit.ly/2t9J7qP)

Put yourself in the shoes of the product team who built the product / feature

* Why do you think they made certain decisions about how it works?
* What could be done to improve the product?
* What kind of metrics you would want to consider when solving for questions around health, growth, or the engagement of a product?
* How would you measure the success of different parts of the product?
* What metrics would you assess when trying to solve business problems related to our products?
* How would you tell if a product is performing well or not?
* How would you set up an experiment to evaluate any new products or improvements?

**Technical Portion (10-20 minutes)**

You will be given 1-2 data processing questions during this portion. This will be a hands-on technical investigation of data problems. We are looking not only for coding skills, but also for the ability to design an operational approach to figure out a concrete answer to a specific question using data.

**Sample Questions:**

* Given timestamps of logins, figure out how many people on Facebook were active all seven days of a week on a mobile phone.
* How do you determine what product in Facebook was used most by the non-employee users for the last quarter? [Required parameters will be given]

While the actual data presented in the technical question will vary based on the question being asked, you can expect to see event-level data, dimension-level data, or both. This data set is designed for interview purposes and is not representative of the large data sets we work with at FB.

**Mock Data and Questions:**

Here's a couple of examples that don't pertain specifically to Facebook but are representative of the kind of data and format with which you'll be presented:

**Event-level data:** an attendance log for every student in a school district

date | student\_id | attendance

**Dimension-level data**: a summary table with demographics for each student in the district

student\_id | school\_id | grade\_level | date\_of\_birth | hometown

Using this data, you could answer questions like the following:

* What was the overall attendance rate for the school district yesterday?
* Which grade level currently has the most students in this school district?
* Which school had the highest attendance rate? The lowest?

You will be expected to write code that would answer the data processing questions given based on a schema or set of schemas that will be provided to you. Whether you choose SQL, Python, or R, please follow standard coding style and best practices for easy readability.

**If using SQL**, you can expect to be assessed on some subset of the following:

* Write a query or set of queries to derive insights based on the given log(s) or schema(s)
* Work with aggregate functions
* Utilize different types of Joins (IE: Left, Inner, Outer, etc.)
* Utilize Union and Union All.
* Work with concepts including Distinct, Random Sampling, De-Duplication, Optimization.
* Apply the results of your analysis to make product decisions or suggestions.

**If using Python:**

* Always check for edge cases and corner cases when you are coding and try to provide a bug free solution.
* Be prepared to improve your solution when asked.
* Make sure to practice the core CS concepts like arrays, linked list, stacks, queues, hash maps, binary trees and graphs, searching and sorting, recursion and parity.

**Python/Libraries:**

* This is a plain text environment, he/she will be on a shared text editor, there will be no libraries available for the interview, this is not a developer's environment. Though it’s not a developer’s environment, candidates will ask about libraries because there's quite a few functions that are only available IF you were to call a library in a real coding environment. So even on a whiteboard or Coderpad interview without execution capabilities, they still want to make sure that they can use functions that are in those libraries when they're writing their code. Most of our Python candidates will almost certainly want to use Pandas, for instance a lot of our R candidates will want to use plyr or dplyr and our interviewers in those languages should all be able to recognize what those are and be able to read the output.
* Interviewers generally know the most widely used libraries, but there's no guarantee that any individual interviewer will be familiar with the libraries you're mentioning. That's okay but you may have to provide more guidance / context around what you're writing or you may be asked to rewrite in base Python.

**If using R:**

* Be thoughtful about implementation efficiency.
* Be familiar with apply functions family, dplyr package or an equivalent.
* While statistical packages can be called directly in R, please also expect some general programming that needs to be implemented, such as data frame manipulation and control flow.

*Technical Resources:*

[SQL Course](http://www.sqlcourse2.com/)

[Programmer Interview SQL Practice Database](http://www.programmerinterview.com/index.php/database-sql/practice-interview-question-1/)

[Mode Analytics SQL Tutorials](https://community.modeanalytics.com/sql/tutorial/introduction-to-sql/)

**GENERAL TIPS FOR THE INTERVIEW:**

* During the video conference interview, you will need access to a computer. You will log in to [www.coderpad.com](http://www.coderpad.com/) - basically like a chat room, so the person on the other side can look at your code. (You will receive additional details about this when the interview is confirmed).
* We strongly encourage you to have a full understanding of being able to analyze data sets through a scripting language of your choice (SQL, R, or Python). If you do NOT know the basics of one of these languages, you should not be moving forward with scheduling an interview. We are happy to give you additional time to prepare and schedule your interview for a few weeks out.
* You can answer the questions in the language that you feel most comfortable with, although choosing a language that is going to assist in getting to an optimal solution in the most speedy and efficient manner is key.
* It's important that you think out loud/ provide a narrative as you go through the problem so the interviewer has insight into your thought process.
* The interviewer will prompt you if you are heading in the wrong direction. Your ability to pivot your answer is a positive signal.
* Feel free to ask clarifying questions during the interview as well.

**ADDITIONAL NOTES:**

* We recognize that candidates may utilize outside sites/resources to prep for our interviews. Please take caution when referring to publicly posted interview questions as we've found many of the proposed solutions to be incorrect.
* Your interview confirmation email will contain the names of all your interviewers (this is subject to change). However, just because an interviewer works on Messenger (for example) does not mean the interviewer will ask you questions related to Messenger.
* During the interview, the interviewers may take notes on their phone or computers. Please do not take this as a sign their attention is elsewhere. They are fully committed to you!
* Please prepare any questions you have before the interview and be prepared to answer why you are interested in Facebook. The interviewers like to see people who know about our environment, projects, challenges, etc.
* If there is time for questions at the end of the interview, please feel free to ask questions focused on Facebook and analytics. Interviewers would be more than happy to answer!
* After the initial interview, I will give you an update regarding next steps as soon as possible. Unfortunately, if things do not work out, I will not be able to share any feedback from your interviews.