

**1) Can you provide an example of when, during a project or analysis, you learned about (or created) a new technique, method, or tool that you hadn't known about previously? What inspired you to learn about this and how were you able to apply it?**

During a personal project aimed at analyzing "coding" playlists on Spotify and creating genre-specific playlists, I explored non-negative matrix factorization (NMF) and scikit-learn's model pipelines.

The goal was to group artists found in the "coding" playlists into coherent genres. I was inspired to delve deeper into NMF and scikit-learn's pipelines because they offered a streamlined approach to preprocess data, create feature representations, and efficiently cluster the data.

By applying NMF and leveraging scikit-learn's pipelines, I successfully identified latent genres within the data and assigned artists to their respective genres. The output of my project can be found in the form of curated genre-specific playlists on a Spotify user profile I created for this project ([source](#)). This experience allowed me to develop a comprehensive understanding of scikit-learn's capabilities and how to use it to implement efficient solutions for intricate data analysis tasks.

**2) A core value proposition of Zelus is packaging and delivering high-end models created by our Data Science team to our clients in an accessible format. Models developed can range in complexity and in the number of data extracts made available to our clients. Describe a project that you have worked on that you feel is most related to building a productionalized pipeline that delivers a model's results to a stakeholder. What were your responsibilities during this project and what tools / techniques did you implement?**

As the creator and maintainer of [JuliaPackages.com](#), I developed a platform that allows the Julia community to discover and evaluate packages more easily. This project involved creating a pipeline to gather package data, process it, and present the results in an accessible format for users.

The project entailed designing and implementing the data extraction and processing pipeline using Julia, web scraping techniques, and APIs to collect package data from multiple sources such as GitHub and the Julia registry. I stayed closely connected with the Julia community throughout this process, gathering feedback and suggestions to enhance the platform's relevance and usability.

To facilitate user access to the data, I built a user-friendly web interface, enabling programmers to browse, search, and filter packages based on their preferences. The website has since attracted over 250k users, with more than 20k monthly visitors. This achievement underscores the successful implementation of a pipeline that effectively delivers valuable information to its users.

**3) Zelus delivers daily updates of proprietary models to their clients, which in some cases contains processes that take 30 minutes to run for a single game on a local development machine for hundreds or thousands of games per day. Describe a project that you have worked on that you feel is most relevant to this challenge and the various techniques you used to accomplish this task. What were your responsibilities during this project?**

At Next Music, I worked on a project to collect information on around 10 million musicians from various web sources, including SoundCloud, YouTube, and Bandsintown. For this project, I was responsible for designing and implementing the web crawling and data extraction pipeline, as well as optimizing its performance.

To handle the vast number of pages that needed to be processed, I employed parallelization and distributed computing techniques using AWS Lambda and other cloud services. This allowed us to scale our solution and process large quantities of data more efficiently. In addition, I optimized the performance of our web crawling pipeline by implementing caching and rate limiting strategies.

We collected data on a per-source basis, processing large amounts of pages for each new data source we added. For instance, we managed to crawl every SoundCloud artist's page in just a week. By overcoming these challenges, we were able to create a comprehensive database of musician information and deliver valuable insights to our team.

**4) Do you have any experience working with cloud computing environments? Please describe example uses of such platforms. If not, what types of projects or scenarios do you think would be most beneficial to deploy a cloud compute environment and what tools might you try to test out or use?**

Yes, I have extensive experience working with cloud computing environments. In my role at Next Music, I worked on a project that involved collecting information on around 10 million musicians from various web sources. As I detailed in my response to question 3, I utilized cloud computing platforms, particularly AWS Lambda, to implement parallelization and distributed computing techniques. This approach allowed us to scale our solution and process large quantities of data more efficiently.

I've also utilized other AWS services such as Amazon S3 for storage, Amazon EC2 for compute instances, and Amazon ElastiCache for managing Redis. These cloud environments have proven to be instrumental for scenarios requiring scalability, flexibility, and cost-efficiency, such as large-scale data processing, machine learning model training and deployment, and running complex simulations.

As for new tools and technologies in the cloud computing sphere, I'm personally interested in platforms that streamline machine learning workflows, like AWS SageMaker and Google's Vertex AI. These platforms offer robust capabilities for building, training, and deploying machine learning models at scale, which is a fascinating area of development in the field of data science.

**5) In your role at Next Music, could you describe the process by which you're assigned tasks or projects? What opportunities do you have to share intermediate results, incorporate feedback from others, and iterate on your work?**

During my time at Next Music, tasks and projects were typically assigned through a combination of recurring team meetings and ongoing communication using collaboration tools like Slack. We had a dedicated team meeting focused on discussing major initiatives and planning for the upcoming week. Outside of this meeting, we used Slack and scheduled additional meetings as needed for discussing smaller updates or addressing more complex tasks.

Our team often collaborated with the marketing department, particularly during larger data meetings, to ensure that our projects were aligned with the company's overall goals. We used project management tools like Notion to keep track of tasks, assignments, and deadlines, which helped us prioritize our work and stay organized.

One of the key aspects of our work process was the continuous feedback loop. We frequently shared intermediate results with our teammates to gather their input and suggestions. This feedback allowed us to make improvements, address any issues, and ensure that the final results met or exceeded expectations. This iterative approach to work helped us maintain high-quality standards and adapt quickly to new information or changing requirements.

**6) In your role at Next Music, what steps do you take to automate, document, or ensure reproducibility of your work? Why do you think these steps are important?**

While working at Next Music, I employed several strategies to automate, document, and ensure the reproducibility of my work. These measures played a crucial role in maintaining consistency, minimizing errors, and fostering a collaborative atmosphere.

- **Data Cataloguing and Documentation:** I made sure that the musician data we gathered was well-organized and documented, with a structured schema and clear descriptions for each data field. This promoted collaboration and knowledge sharing within the team.
- **Version Control:** I utilized Git for version control, which helped us track our work history and easily revert to previous versions if needed.
- **Automation:** I implemented scripts and tools like cron jobs to automate certain processes, which saved time and allowed us to focus on higher-level tasks and problem-solving.
- **Data Ingestion Scripts:** I created scripts for ingesting data from various sources into our Snowflake data warehouse, simplifying our data processing pipeline.
- **Jupyter Notebooks:** I relied on Jupyter Notebooks for data analysis and visualization, documenting my process and findings to facilitate sharing my work with colleagues.

Implementing these measures was essential as it helped us maintain high-quality standards and create a collaborative environment. By automating, documenting, and ensuring reproducibility, we could easily share knowledge, learn from one another, and build on our collective expertise.

**7) Can you provide an example of a paper, presentation, or article relating to sports analytics that you found particularly interesting?**

One example of an article relating to sports analytics that I found particularly interesting is "How Many Sources Do You Need to Consistently Cast Your Spells?" by Frank Karsten ([source](#)). The article focuses on Magic: The Gathering, a popular card game that has also become an esport.

In this paper, Karsten effectively uses data analytics to guide players on building their decks based on statistics, which is a fundamental aspect of the game. He examines the probabilities of drawing the right cards, helping players make informed decisions about their deck composition.

Websites like 17lands.com and untapped.gg further demonstrate the application of analytics in the game, as they provide various success metrics for every card. This information is invaluable for players looking to optimize their gameplay and make strategic choices based on data-driven insights.

The application of data analytics in Magic: The Gathering and esports showcases the potential of data-driven approaches to improve performance and decision-making, which can also be translated into the realm of traditional sports, further enhancing players' and teams' understanding of their respective games.

**8) What experience do you have working remotely? What do you see being the biggest advantage of remote (or asynchronous) work? What do you see as the biggest challenge?**

I have been working remotely since 2019, initially while collaborating on MIT's reactor project from NYU, and later as a data engineer at Next Music. Throughout this period, I've gained valuable experience and insights into the advantages of remote work.

Remote work offers several benefits, including the opportunity to carefully consider and formulate responses before communicating, which can lead to clearer and more thoughtful communication. Additionally, remote work provides flexibility in terms of work hours and location, which can contribute to increased productivity as it allows employees to work during their most efficient and focused hours. This flexibility can also improve work-life balance and overall job satisfaction, ultimately leading to a more motivated and engaged workforce.

One challenge of remote work is coordinating schedules across different time zones. However, I've found that with proactive communication and being mindful of team members' availability, it's possible to find mutually convenient times for meetings and collaboration.

Overall, I believe that the advantages of remote work far outweigh the challenges, and I'm confident in my ability to navigate these challenges effectively to ensure a productive and collaborative work environment.