The project I have been working on since I started at ZS involves monthly processing of large quantities of sales data (~billions of records) from IQVIA. This data, combined with various mapping files, is processed and reports are sent out to the client, a prominent pharmaceutical company. The processing involves cleaning the data, performing quality checks, merging with rep alignment data to credit sales to sales teams, and roll ups to multiple higher geography levels and business segments. Our project’s outputs and reports drive the company’s sales force management and incentive compensation functionalities.

All processing is done using a framework that leverages AWS s3 for data storage and Redshift as a database, running automated Shell scripts to fire PostgreSQL queries on the data in order to perform various transformations, merges, and quality checks. My responsibilities involve developing and automating data quality checks, preparing system configurations, and running end-to-end system processes. At various stages during our processing, reports go out to the client, both to validate input data, as well as to highlight any other observations or changes we might find. What makes this project interesting for me is that the scale of the data and processing involved requires an in-depth knowledge of the system, as the smallest of errors in data can result in a cascaded impact that affects all cross-work streams.

I have also developed Shell scripts to automate certain data transfer or data quality management processes- this has saved a considerable amount of processing time. Writing and testing these scripts not only demonstrates my algorithmic and programming abilities, it ensures I have the computer science background required to pursue an analytics career.

Working on this project has enabled me to develop a robust analytical mindset- one of the most interesting things I get to work on is performing impact analyses based on data issues (such as discontinuities in data reporting) reported by IQVIA. Each issue being different from the next, the variety in the nature and extent of impact always varies. Learning and navigating the intricacies of a system as vast as the one I work with, along with performing such varied impact analyses, has honed my analytical skill. Furthermore, the varied nature of issues has always required a renewed approach to solving each one, and I have been able to enhance my problem solving skill significantly.

These analyses go directly to the client, and on the basis of the extent of impact, various adjustments to the data we provide need to be made. The choice of how to represent the analyses- graphs, Excel reports, or a simple e-mail to the client- is also one that is left with us, and has enabled me to significantly improve the quality of my written communication. All of these factors combined are what make this project interesting for me- there is always something new to learn, and always opportunity for growth.