

TO: Allen Renear

FROM: Harsiddh Patel

DATE: November 14, 2021

SUBJECT: Importance of Data Curation

Dear Mr. Renear,

I would like to welcome you to our Data Science department. With your presences, the department will have tremendous growth.

I am writing this to inform you the importance of data Curation to our data science projects and will be going over some of activities of data Curation. To begin with, data Science in general is collecting data for analysis, cleaning, aggregating, and manipulating the data to perform advanced data analysis. In this process, data curation is indirectly part of data science since it provides data integrity and usability to do future development or research. These are important for our company since data accuracy and analyses are paramount. We will go over three goals of data curation; Provenance and Preservation.

Provenance is crucial for any data science project since we use data from multiple sources to make next decision for our company and for their clients. Provenance ensures that any input of data to a system is well documented at every stage it goes through. So, in the times of crucial downtown of an application, developers can easily pinpoint the cause of the downtown by going through the documentation. In addition, provenance creates trust in the report since it provides transparency to the system which will in turn create trust to our clients.

Preservation is the goal that ensures that data is easily understandable and the convenient of useability/reusability in the future. This is the idea of data curatorial activities. The documentation of application schema, metadata, and the all the policies needs to be maintained for long-term preservation. If our company don't invest in this today, then it will cost more in the future and give out bad reputation of our company to clients.

In conclusion, having data curation services is crucial to our company since it provides benefits that are not clearly visible but without it, the company would suffer. The funding to data curation services should not stopped but should prioritize since data is an asset to our company and we should take every step to ensure it is well protected.

Sincerely,

Harsiddh Patel