Innovative Data Solutions Case Study

Extracting Data from PDF and HTML Reports

Challenge and Opportunity

A large insurance company had hundreds of thousands of forms which underwriters had to manually read to approve applications. The name of the game in insurance is getting an accurate quote first.

Approach

- Use Python to parse reports and extract required data.
- Save all data to an Exasol database.
- Maintain detailed entries on the several types of errors encountered.

Results

The system was able to parse 215,273 and found 57,529 valid issues with the reports that the client was previously unaware of. Aside from the automated QA, we are able to enhance/clean the data through NLP methods. This improved speed and accuracy of the previous method to underwrite these policies.

50+ **Clients Served***

20+ **Data Services Offered**

185% **Average ROI**
Based on 2 years of cost decrease or revenue increase over consulting fees *

(excluding internal implementation cost)

Keiter
Your Opportunity Advisors

*Engagements performed by team members prior to joining Keiter CPAs