

Ayush Maheshwari, Ishita Shah, Richa Tiwari Choubey, Matthew A. Lanham
 Purdue University, Krannert School of Management
 mahesh15@purdue.edu; shah567@purdue.edu; rchoubey@purdue.edu; lanhamm@purdue.edu

ABSTRACT

No one size fits all when it comes to health insurance, and the need of the hour is to build customized insurance plans that would reduce expenses of organizations towards providing health benefits to the employees. To create these tailored insurance plans, and provide the best option, one of the key ingredients is to have the correct data.

INTRODUCTION

The client currently uses a manual approach to extract important census metrics that are directly important to prioritize counties for sales calls. Through this process, they extracted data from census.gov for 30+ states and 2500+ counties. The overall process takes 30 hours every week for 2 months which amounts to total work of 480 hours. Through process automation, our team has reduced this time to less than 10 hours resulting into reduction by 97.9%. We embarked on this journey of building these automated data extraction pipelines using python, which scraped the US census website and stored county-level demographics. We built Tableau dashboards that would provide insights to the sales team on creating these tailored insurance plans for the organization. This data extraction process was manual and not scalable. Therefore, our team went on to automate it, eliminate susceptibility to failures and reduce data quality gaps. This new data extraction process is less susceptible to data quality gaps, failures and the more importantly takes less than 1 business day to be completed which is a 98% savings of man hours.

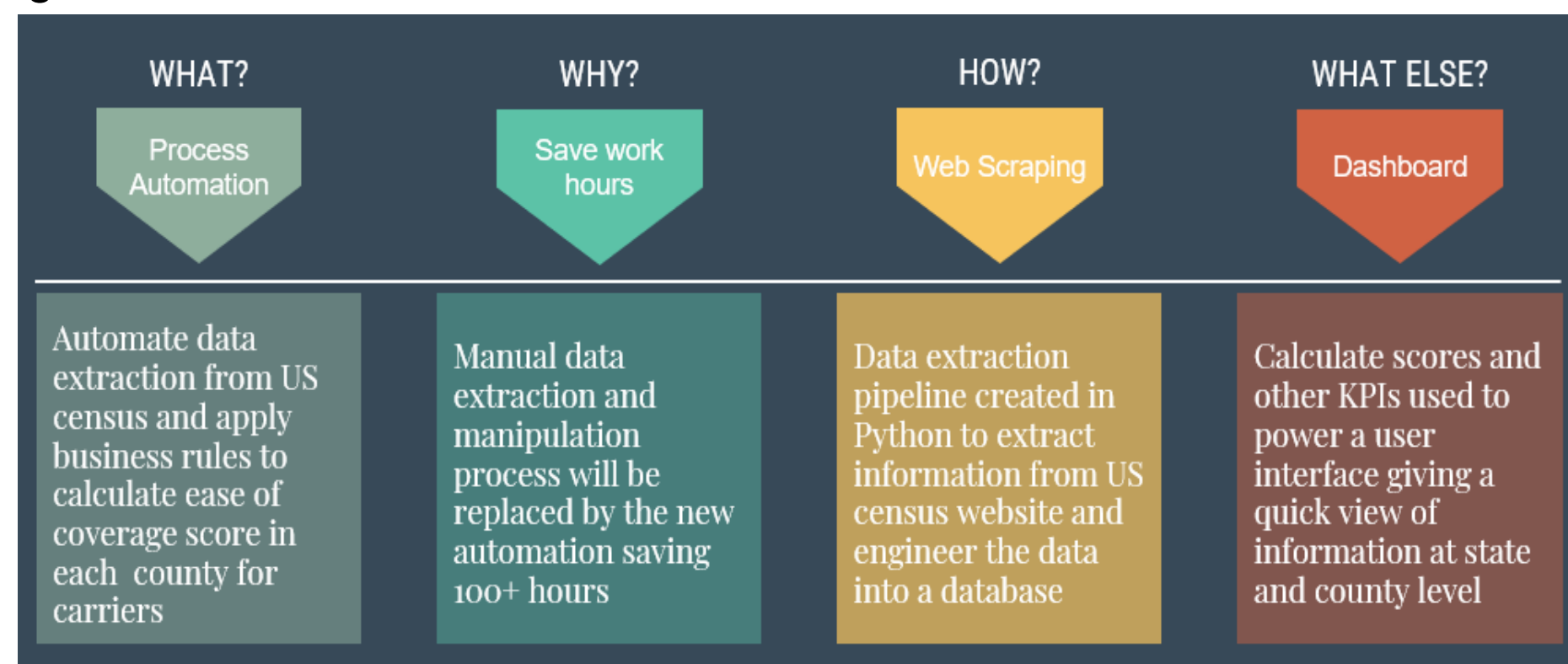


Fig 1. Overview of Business Benefits

RESEARCH OBJECTIVES

- Minimize human intervention needed in the data collection process
- Reduce the time required to collect data and store it in client's database
- Creating a dashboard to show counties that should be prioritized for sales calls

LITERATURE REVIEW

Study	Techniques	Time	Memory	Data Usage
Will Koehrsen, 2018	Regular Expression		✓	
Neal Haddaway, 2015	HTML DOM	✓		
Richard Penman, 2019	X Path			✓
Our Study, 2021	Regular Expression		✓	✓

METHODOLOGY

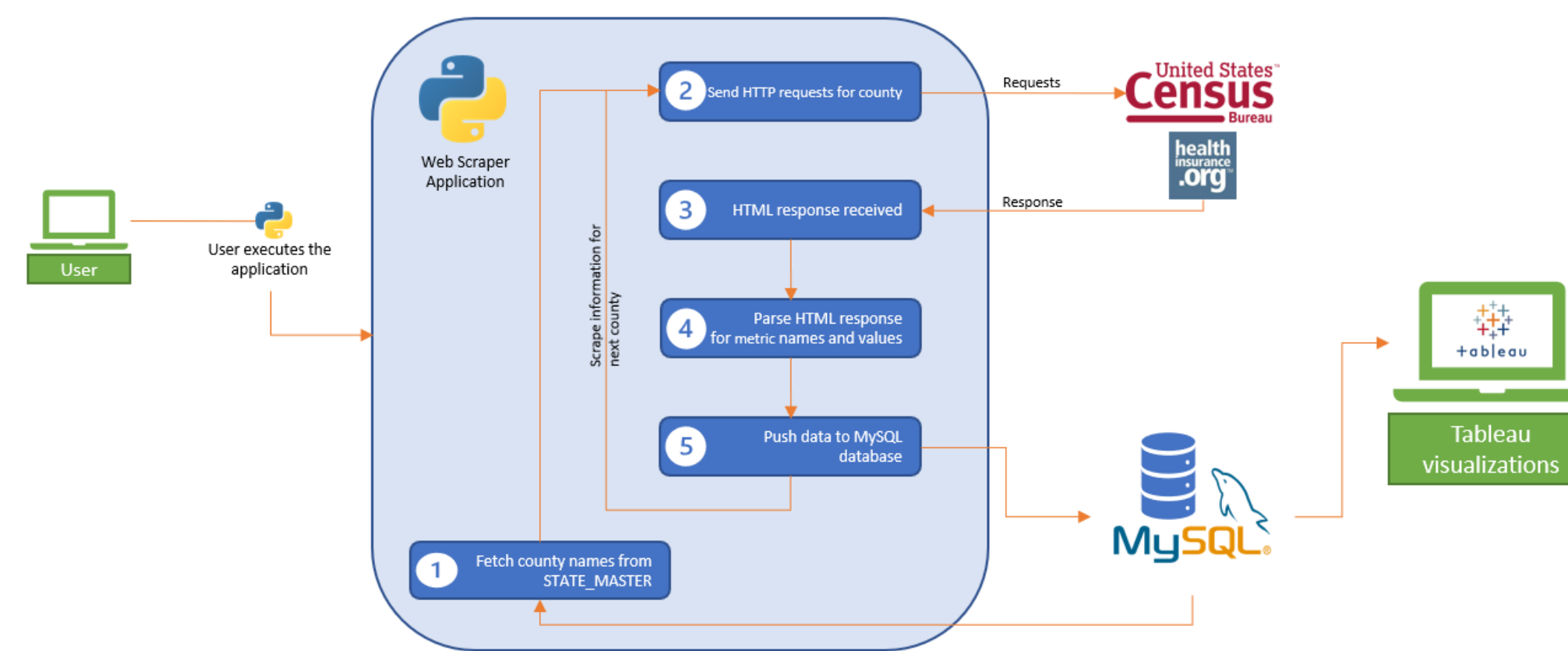
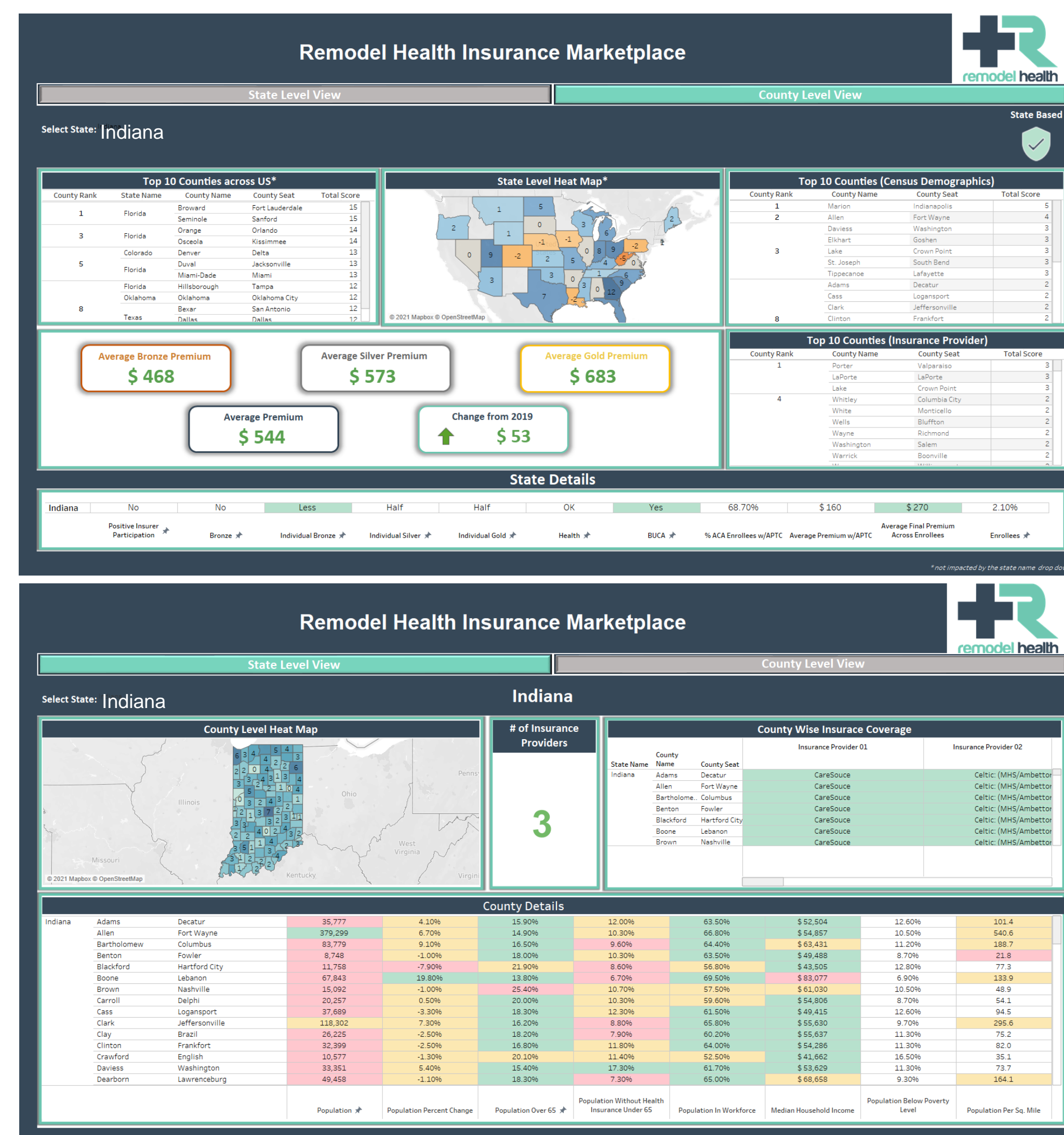


Fig 2. System Design

List of counties fetched and fed it into the scraping application. Data extracted for each county stored in MySQL and business rules applied on the metrics to generate insights on insurance coverage across counties. Feeding this data to Tableau helps in effectively effectuating insights on insurance premiums across states.

DASHBOARD OUTPUT



BUSINESS IMPACT

Manual data collection required allocation of two associates to this activity for a period of two months. These associates would spend 30 hours every week, resulting into a total of 480 hours needed for the entire data collection processing. On replacing the manual processing by an automated web scraping script, the total data collection time has reduced to 10 hours, resulting into saving up to 98% of time, thus freeing up resources for other work.



As the automation process reduced the need for human intervention, consequently it minimized the scope of human error. Our team identified several typing errors in county names and in rules applied for calculating the potential of each territory. This saved the client from missing out on potential sales opportunities.

If population change < - 5% then, use red highlight for the cell

County	Population Percent Change
Blackford	-7.90%
Fountain	-5.20%
Grant	-6.10%
Pulaski	-7.70%

Business rule

Incorrect color coding



The Tableau dashboard had a direct impact on the efficiency of the sales teams to conduct business. The geospatial design with color key gives a brief overview of which areas within a state have higher business potential, thus helping the regional sales managers in resource allocation and creating revenue estimates. Additionally, the rankings of top 10 counties based on their coverage score help the sales personnel to prioritize sales calls that have higher business potential.

KEY TAKEAWAYS

Remodel Health's business problem may not seem complicated at first, but on looking closely we can see how automating any manual process can have far reaching effects on the business by –

- Saving man hours
- Freeing up resources
- Minimizing human errors
- Avoiding potential business loss by correct identification of sales opportunities

ACKNOWLEDGEMENTS

We would like to thank Remodel Health team and Professor Matthew Lanham for their guidance and support on this project.