

<INDUSTRY SURVEY ON WEB SCRAPING/>

Table of Contents

| | |
|---|----|
| Introduction | 3 |
| Different approaches to data acquisition | 3 |
| The data ecosystem | 3 |
| Insights on Web Data | 4 |
| Major roadblocks in the web scraping industry | 5 |
| Insights on the industry | 6 |
| 1. Business needs for web data | 6 |
| 2. Budget for data acquisition..... | 6 |
| 3. How data helps | 7 |
| 4. Effectiveness of data depends on support from management | 8 |
| 5. Business needs and success map of web tools needs | 8 |
| 6. Critical drivers of web data usage..... | 9 |
| Business-Technology mapping..... | 10 |
| Target market profiling | 11 |
| Top challenges | 12 |
| Conclusion..... | 14 |

Abstract

The adoption of web data is on the rise among major industry verticals as it provides deep insights on the business and its competition. This power of web data makes it a resourceful tool to improve the overall efficiency of the organization while also boosting the revenue in the process. Statistics show that the usage of data in business is steadily increasing. As the data available on the web keeps growing by each passing second, the web data extraction industry is expected to reach new heights in the near future.

Introduction

The web scraping industry is growing steadily with the increasing demand for web data acquisition by companies across industry verticals. Businesses have realized the potential benefits that can be harvested from the unstructured data spread all over the internet. This is one of the key reasons for the proliferation of companies venturing into data acquisition space. As the competition gets tighter by each passing day among businesses, the need for generating insights and making data-backed decisions is becoming crucial. To evaluate this growing market of web scraping and data acquisition solutions, we conducted an in-depth survey. This is a compilation of our findings from the survey on web scraping industry.

The need for data stems from the increasing competition which makes it a bad idea to make decisions based on mere assumptions and instincts. Having solid data that can help you analyse the situation from a neutral viewpoint helps in this scenario. Since the web is an ever-growing repository of data, companies are now venturing into acquiring this data using web scraping technologies. One factor that makes web scraping a service that should be outsourced is its complexity. Web scraping is a niche that is technically challenging. The dynamic and ever changing nature of websites on the internet also adds up another complexity - monitoring the sources. Companies want to focus on their core business rather than acquire data, as the challenges mentioned above can be overwhelming for businesses without any expertise in data extraction. Considering this, the web scraping industry is set to grow many folds from here. Depending on a web scraping service provider is the best bet for enterprises as they can get the data they need for business intelligence without worrying about its acquisition.

Different approaches to data acquisition

Web data can be acquired by various routes. Here are the most common ways by which data can be acquired for business intelligence and other similar use cases:

- Web Scraping service (Preferred by enterprises)
- In-house web scraping (Not an ideal solution; used by some small businesses)
- Web scraping tools (Not suitable for enterprise-grade data acquisition)

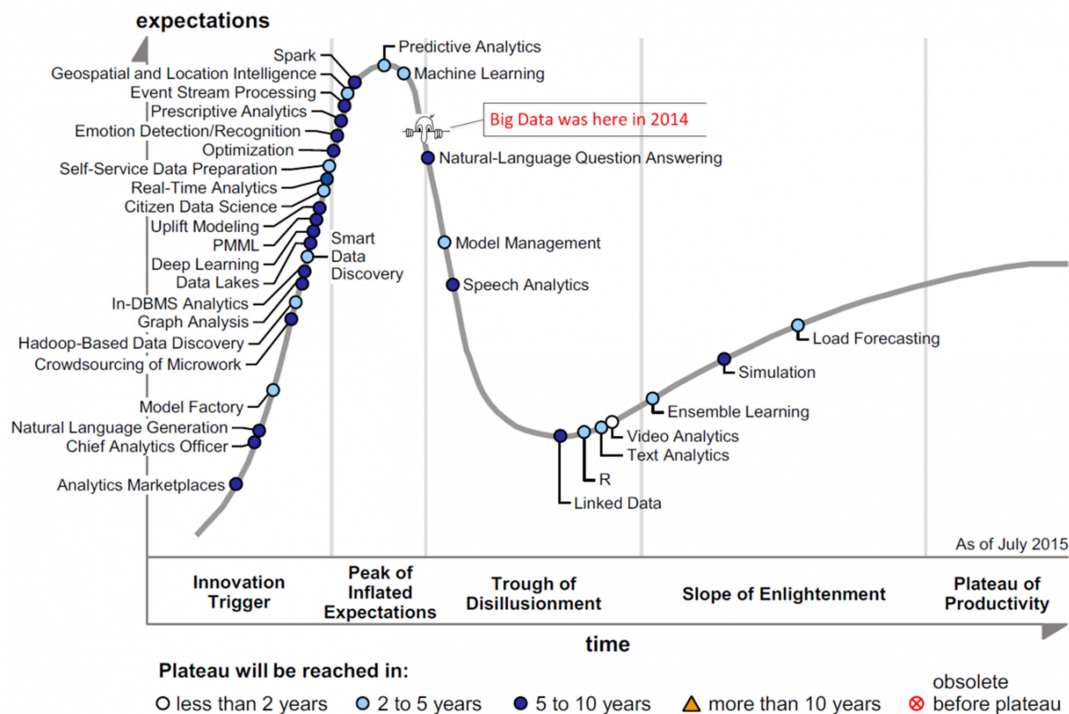
Web scraping is clearly the best solution when it comes to getting data for a serious business purpose.

The data ecosystem

Machine learning and predictive analytics have already reached a point where they are being applied to complex analytics systems and artificial intelligence applications. As these applications improve over time, the demand of more data will also increase which looks promising for the web scraping industry. Given below is a report by Gartner that shows the evolution of big data and some of its

complex applications.

Figure 1. Hype Cycle for Advanced Analytics and Data Science, 2015



Source: Gartner (July 2015)

Insights on Web Data

There are many companies that are yet to embrace big data and its real-world applications. While some of them are not aware of the ROI of web scraping, some are facing minor roadblocks; in essence acting as barrier for data acquisition.

Lack of awareness of possible use cases of web scraping

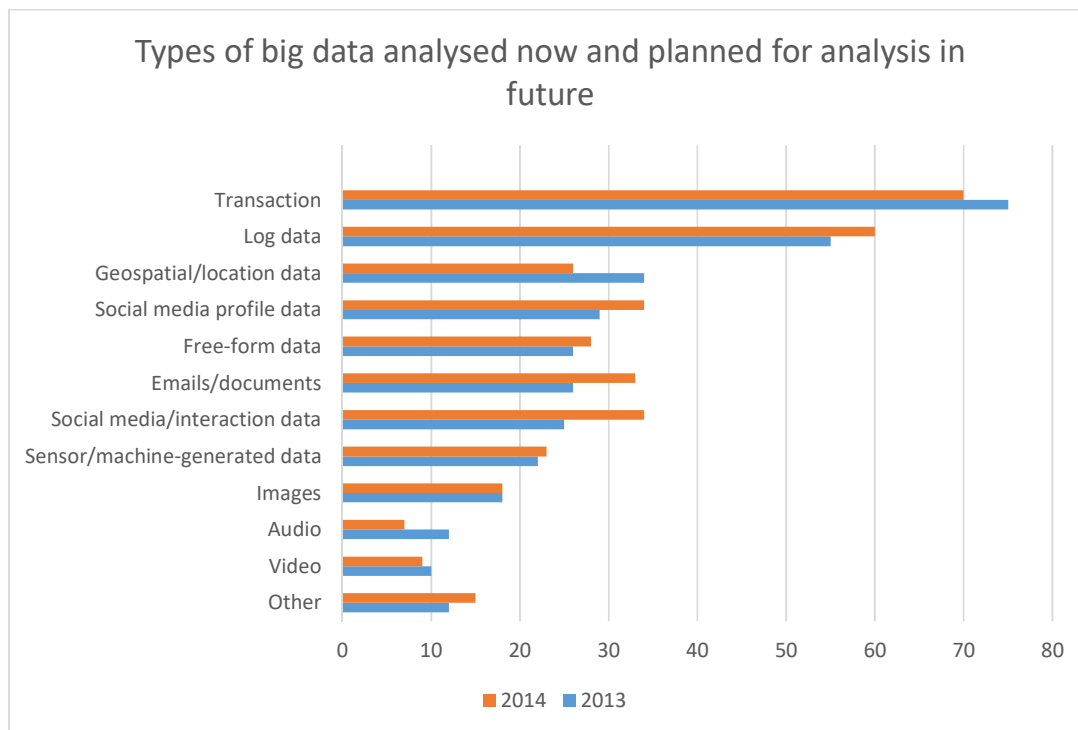
Conceptualizing semantics are critical to the understanding and interpretation of text data and they require participation and agreement across multiple parties.

Ever evolving big data needs of customers poses major problems to the web scraping market.

Intense competition in the market and robust security measures by companies

Web scraping has evolved into identifying concepts, themes and relationships.

The chart given below shows the types of big data that was being analyzed in 2013 versus 2014. It's clear that the amount of data being analyzed has increased for most types of data in just one year.



Major roadblocks in the web scraping industry

The web scraping industry is in its growing phase, meaning it still has to overcome a few barriers and challenges. As the requirement for data keeps increasing, the technologies involved must also evolve in order to cater to the rising demand. Five main factors keep web scraping from becoming the powerful solution to many needs that users and vendors wish for.

- Immature technologies and products
- Difficulty in fitting into a larger ecosystem
- Diverse use cases
- Short supply of sophisticated skills both in IT organization and the end user community
- Web scraping tools is not a single market
- Despite some consolidation, the data mining and predictive analytics market is still highly fragmented, and its diversity complicates the planning and buying process.

By implementing smarter techniques like machine learning and predictive analytics to the existing system, web scraping industry can easily overcome these hurdles and be more efficient and powerful. If machines can intelligently identify data points based on their content, the manual work involved in the setup can be considerably reduced. As the market keeps growing, it is safe to assume that talent required to take care of the process will also be adequate in supply. More use cases of big data are still being discovered by companies which will result in a higher demand for data and data extraction services.

Insights on the industry

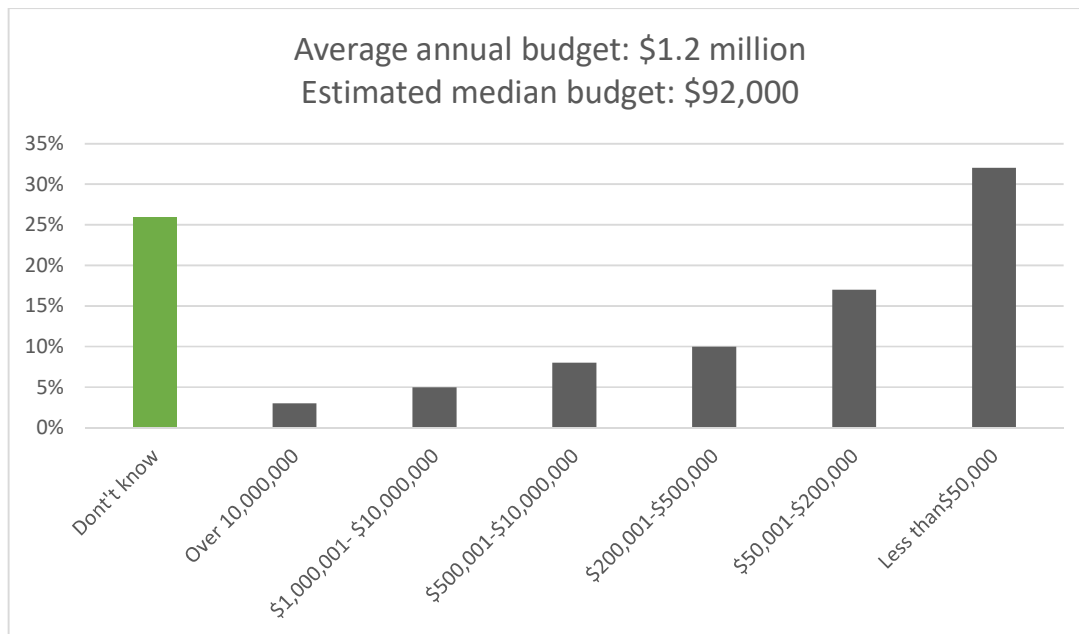
1. Business needs for web data

Web data is most frequently used for financial and market research, sales intelligence and lead lists, competitive and pricing intelligence along with customer sentiment analysis. It is expected that usage of web data in other fields will also grow in the near future and new use cases will come into the picture.



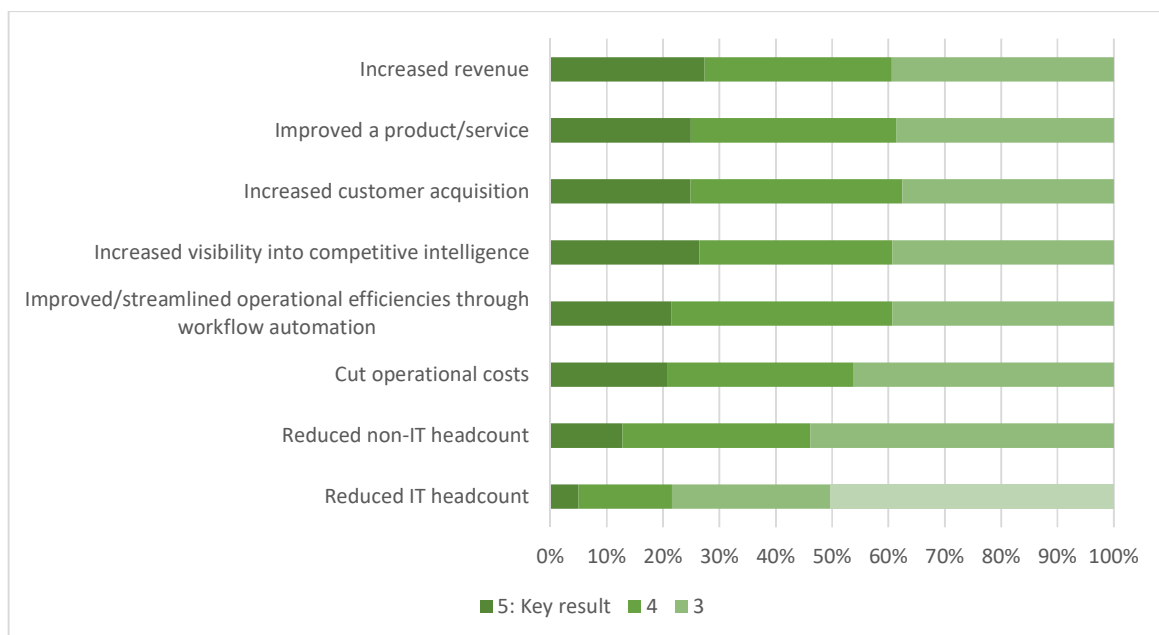
2. Budget for data acquisition

Most organizations appear to have annual budgets for web data collection, monitoring and analytics that is \$500,000 or less. 26% of the organizations don't have a budget decided for data acquisition meaning data acquisition is not under their radar. Not understanding and utilising the power of data can be fatal in today's highly competitive business scene. Organizations that added data into their business strategy have had significant improvement in terms of revenue and customer satisfaction.



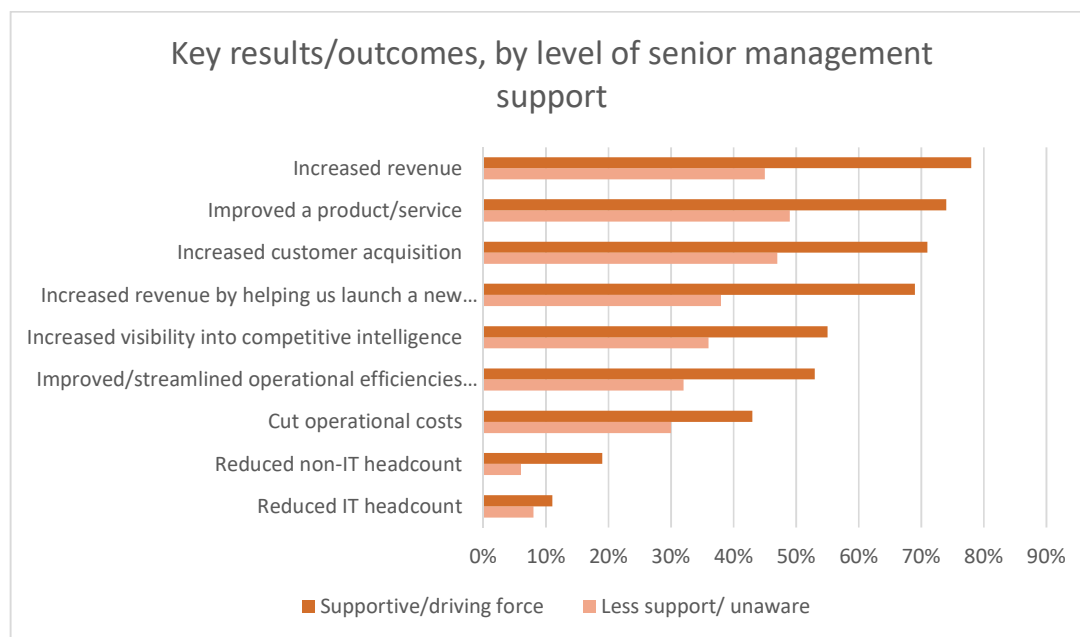
3. How data helps

Web data programs are most likely to yield results by increasing revenue, improving products or services, increasing customer acquisition and launching new products. Given the power of data to improve all aspects of a business by helping the management make better, informed decisions, this shouldn't come as a surprise. Since data backed decisions are not likely to go wrong, the improvement affects all areas of the business positively. Companies that rely on data for their key operations are also likely to use data for business development and competitive intelligence.



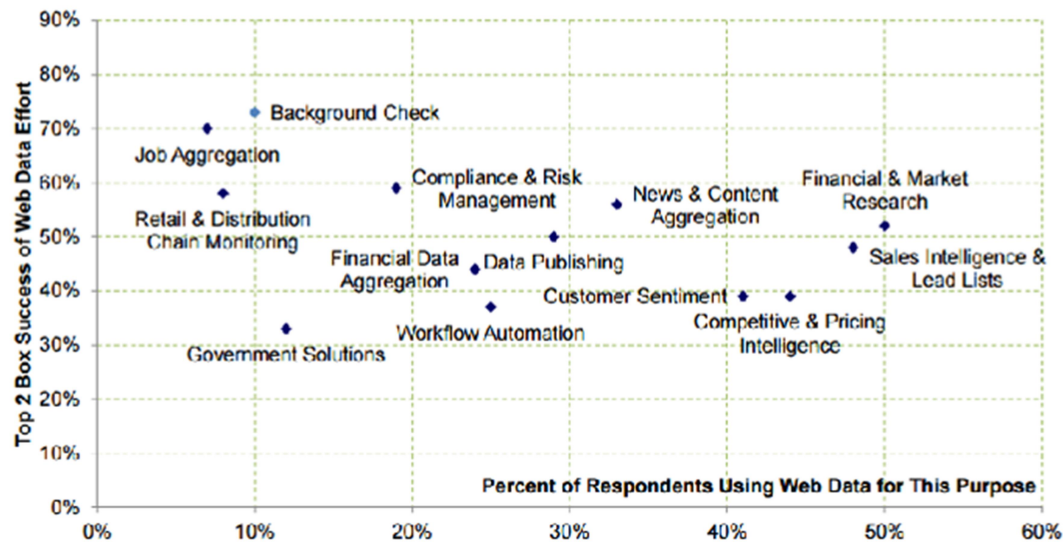
4. Effectiveness of data depends on support from management

Strong support from senior management correlates with a greater number and variety of business outcomes expected from web data. Since data backed business is still a novel concept, the support and encouragement from management is not adequate in many organizations. This support decides the effectiveness of the whole process since the implementation of ideas based on data requires approval from the top management of the organization. When data is given priority by the management, the results seem to be 2x better. This shows the need for better awareness about big data and its power to transform a business from ordinary to extraordinary.



5. Business needs and success map of web tools needs

Web data has use cases in every industry vertical, but the adoption of data seems to be mainly concentrated on a few industries alone. Finance, sales, competitive intelligence and customer sentiment analysis are some of the most common use cases organizations seem to use data for. The success rates are also higher in these areas. This could be because of the interrelationship between these use cases and the web. Financial and market research gets the highest returns from investing in web data.



6. Critical drivers of web data usage

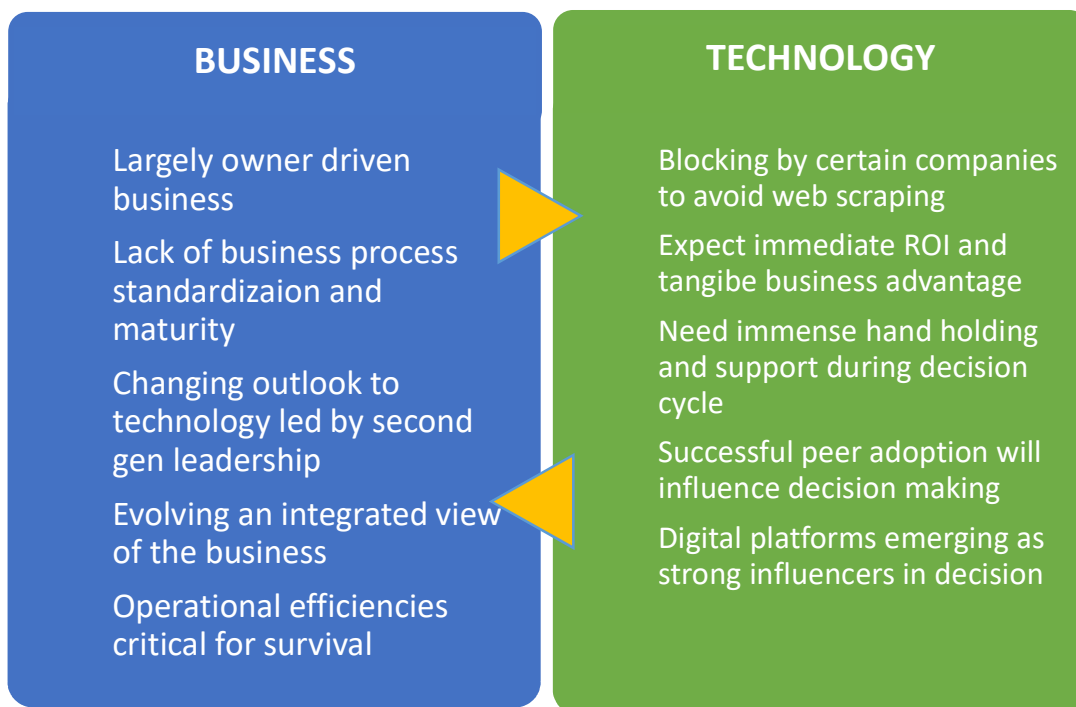
The evolution of means for data extraction is an ongoing process. As the technology keeps getting advanced, the ease of web data acquisition also gets a boost. For most organizations, the success of the current data projects happens to be the motivation to improve existing systems of web data acquisition. Revision of budgets to accommodate more funds for data acquisition is also something that helps in this evolution. Web data use is definitely on the rise across most industries although some of them have a lot of catching up to do.

| Critical Drivers (Top 2 Box) | Job Function | | | | Industry Sector | | | |
|---------------------------------|--------------|------------------|-----------|----------------------------|------------------------------------|-------------------------------------|--------------------------------|------------------------------|
| | Total | General Mgmt. | Marketing | Advertising / Marketing | Business Services Consulting | Financial Services/ Insurance | Publishing / Info/ Media | Software/ Techno- logy |
| Business needs | 73% | 76% | 77% | 80% | 59% | 63% | 82% | 79% |
| Success of current projects | 68% | 80%↑ | 68% | 67% | 55% | 67% | 59% | 89%↑ |
| ROI | 68% | 68% | 77% | 73% | 68% | 58% | 76% | 74% |
| Budget | 52% | 60% | 74%↑ | 47% | 55% | 58% | 71% | 47% |
| Senior management direction | 57% | 64% | 55% | 60% | 59% | 42% | 71% | 47% |
| Market demand | 49% | 60% | 48% | 73%↑ | 64% | 38% | 35% | 68%↑ |
| Technology advancements | 49% | 64%↑ | 52% | 67% | 41% | 42% | 47% | 63% |
| More competitive landscape | 47% | 32% | 61%↑ | 33% | 45% | 42% | 35% | 47% |
| Sample Size | 151 | 25 | 31 | 15 | 22 | 24 | 17 | 19 |

Q10. On a scale of 1 to 5, what will be the critical drivers in the evolution of your Web Data usage in the next 18-24 months? (n=151)
Data reported here for most common job functions and industry sectors with at least 15 survey respondents. Notable differences indicated by ↑.

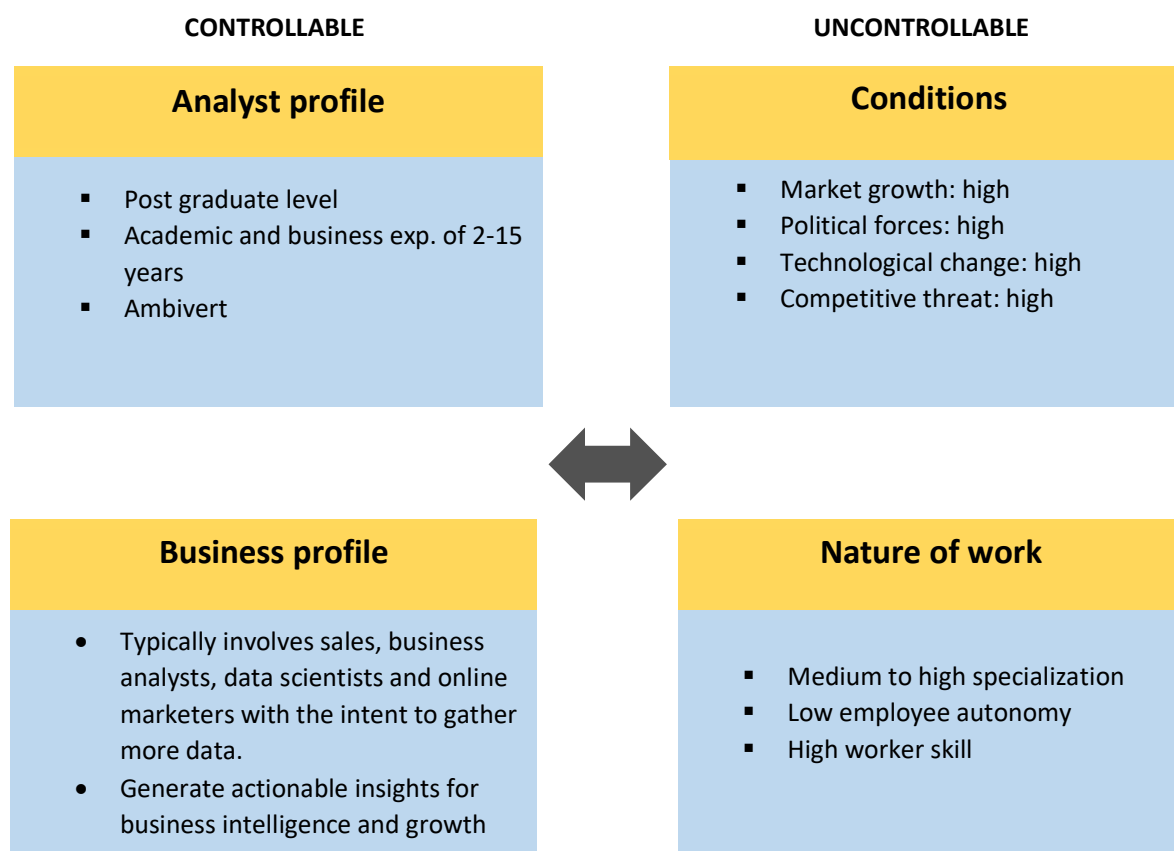
Business-Technology mapping

The integration of technology and its advantages like large scale data extraction is what makes today's data-backed businesses successful. Old style businesses and technology companies both have a different outlook towards web data. While some organizations have successfully integrated data in their core strategy, some are lagging behind due to various reasons. Here is an illustration of the Business - Technology mapping.



Target market profiling

The below illustration shows the target market of web data. It's clear that web data acquisition is an industry that will have to rapidly adapt to various changes in technology, market fluctuation and external factors. Data acquisition and handling are areas that need specialization and high level skills.



Industry verticals:

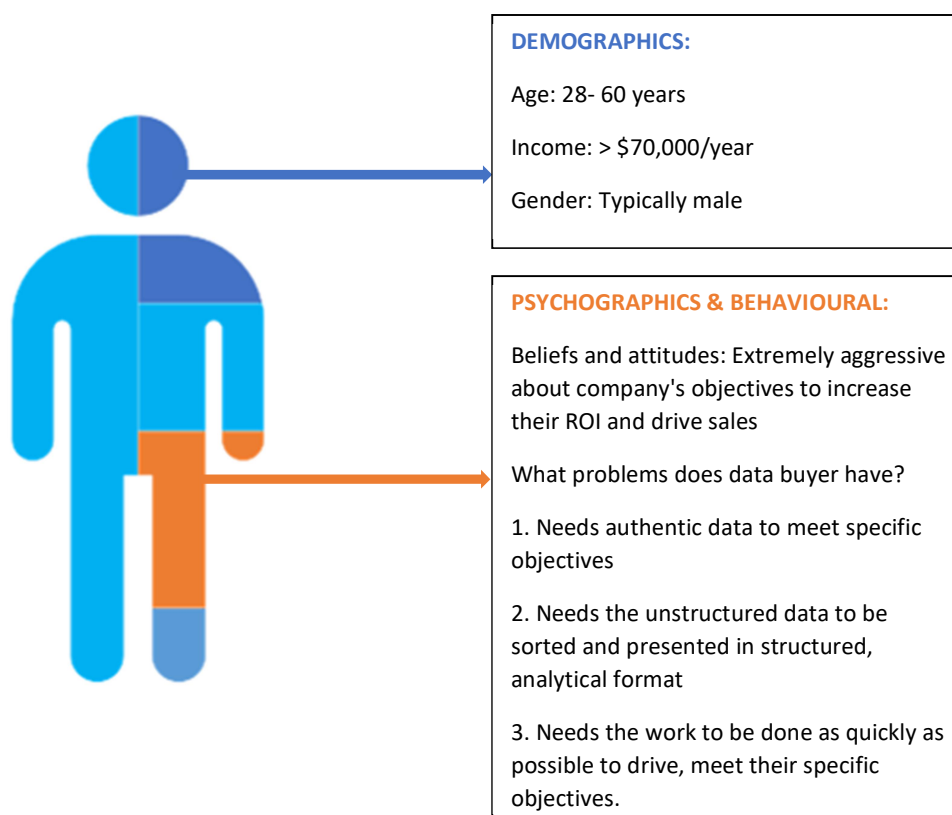
Lead generation: Companies use it to scrape the email and phone numbers

Recruitment: Recruiters use it to collect data in real time to provide live tracking details.

Travel companies: Collects trending topics and use hashtag to collect information from social media profiles.

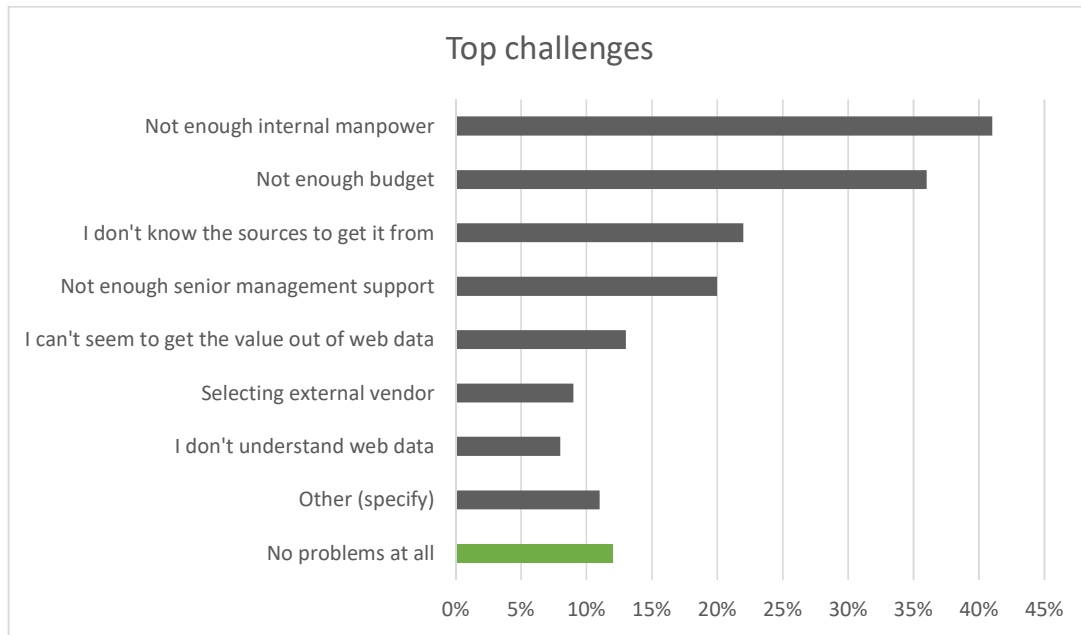
Business directories: Scraping completes information about the business profile, address, email, phone, products/services, working hours, geo codes, etc.

Each business has competition in the present world, so companies scrape their competitor information regularly to monitor the movements. This monitoring has become a vital part of the business activities of organizations as the competition has reached the next level from the implementation of data. Failure to acquire and utilize data can be a huge misjudgment given the intensity of competition among various industries.

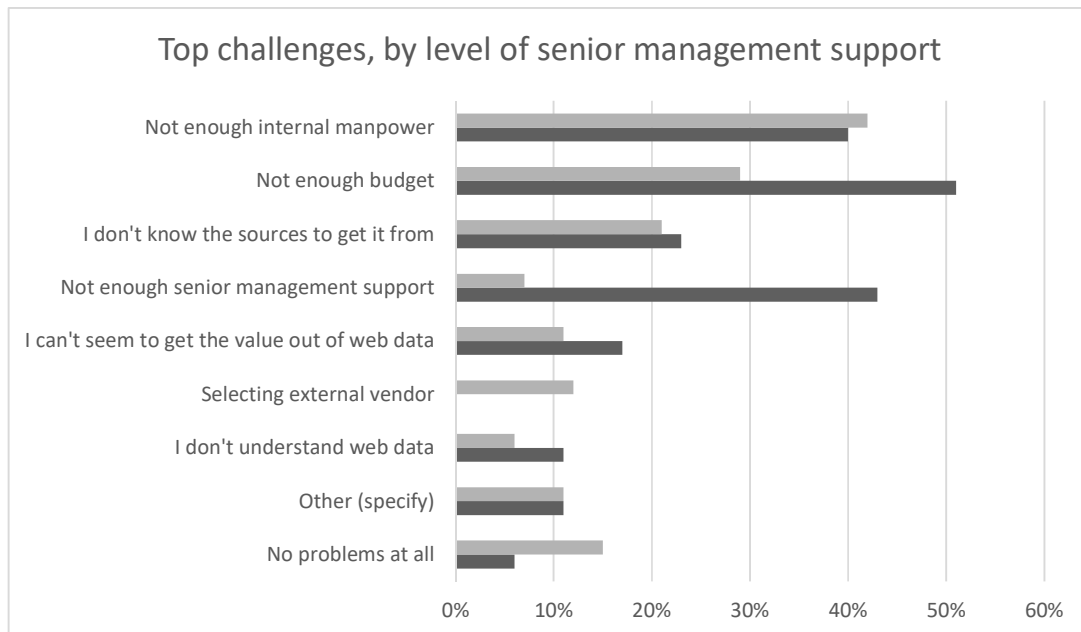


Top challenges

The top challenges faced by companies using web data are a lack of internal manpower and inadequate budget, especially for those in marketing roles. Since web data extraction requires technical skills, the lack of talent itself is the biggest challenge. Organizations also seem to have thin budgets for web data acquisition making it further difficult to acquire data. The primary reason for low budgets is the lack of proper awareness about the benefits of data. Not knowing the suitable sources for data extraction and lack of support from the senior management are also leading challenges when it comes to web data.



In companies where support from senior management is low or non-existent, not enough budget is the leading reason for low adoption rates.



Conclusion

With the competition in the market increasing by each passing day, the demand for high quality data is on the rise. Since business decisions made without solid data could be simply assumptions, they can be fatal to the organization. As more companies have started venturing into data acquisition, it is clear that the web data market is growing steadily. Once the technology barrier comes down and more companies become aware of the need of data-backed business decisions, the industry will experience rapid growth and face equally complex challenges.

Need more information on the subject?

Contact us:

info@promptcloud.com

www.promptcloud.com