LET'S TALK DIGITAL

OCTOBER 2020

ASIAN BANKING SCHOOL

CONTENT

Let's Talk Digital is a monthly newsletter that was created to build awareness on Digital Banking and provide a platform for industry practitioners to share insight and current trends on this exciting subject matter in relation to the Banking and Finance industry.

Applications and a Guide to Web Scrapping

By Koh Wyhow

All organisations have some internal data, whether transaction or customer records. They are also aware of external data all over the internet, from the Department of Statistics Malaysia, or public Application Programming Interfaces (APIs). This article explains a basic approach to web-scraping to enrich internal data sources, which can be used to achieve business outcomes.

Alternative Data & Credit Scoring for the Unbanked

By Peter Kua Seng Choy

Both consumers and financial institutions can now benefit from alternative data for credit scoring. Customers with insufficient credit records would have the opportunity to obtain loans and lenders would be able to apply alternative data to significantly strengthen their credit risk modeling.

To find out more about the Digital Banking programmes that ABS offers, visit

www.asianbankingschool.com/our-programmes/centre-for-digital-banking

Koh Wyhow

Koh Wyhow is the manager of the data science team at Star Media Group Berhad. He focuses on delivering advanced analytics and business intelligence solutions for the organisation like chatbots and image recognition solutions. He consulted for client in the airlines, media, property, and FMCG industries during his time as a senior consultant at EY's Data and Analytics team.

He was one of the data scientists which implemented strategies to run a national data-driven campaign for INVOKE in the 14th General Elections. As an independent learner, he picked up basic Python programming skills after office hours during his days as a Further Mathematics lecturer at a private college. Wyhow holds a BSc in Mathematics from the National University of Singapore.

Peter Kua Seng Choy

Peter Kua is currently Head of Data Science and Analytics in Media Prima Digital. His responsibilities include finding ways data can be used as a competitive advantage as well as identifying new business opportunities with data.

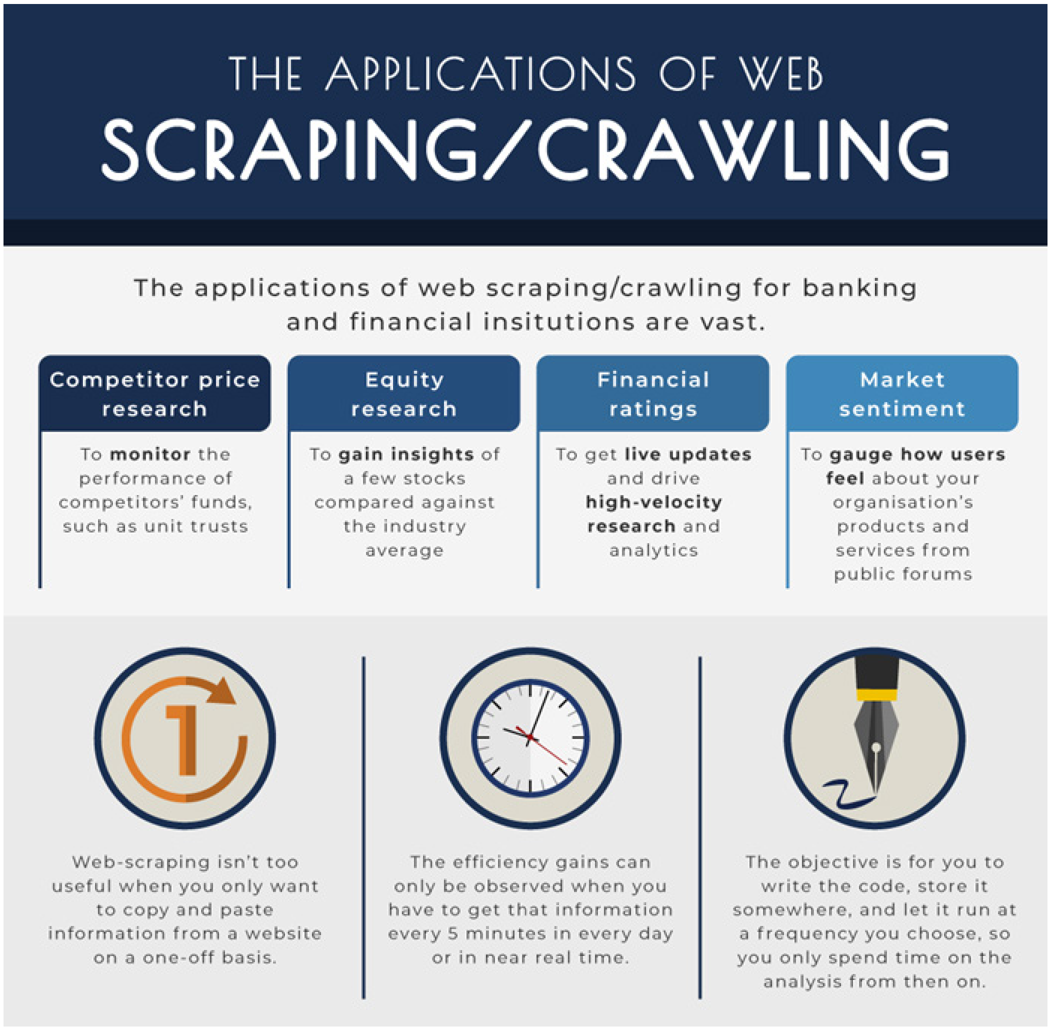
Peter was also instrumental in driving the National Big Data Analytics (BDA) Initiative under the Malaysia Digital Economy Corporation (MDEC) in the areas of thought leadership and industry development. He played a key role in developing the first National BDA Framework that delivered strategic recommendations / action plans to achieve the National BDA vision.

Peter has extensive tech-related experience in various roles: Big Data / Data Science Strategy, Technopreneur, CTO, Project Manager and Software Developer. Startup leadership & management style. Excellent communication skills. Solid network of contacts in the private sector, government and universities/colleges.

Applications and a Guide to *Web Scrapping*

By Koh Wyhow

This article is about making use of external data to complement your organisation's data sources by web-scraping. Most organisations like financial institutions have rich datasets on their users: they are able to identify who are the large spenders, who are likely to register for exclusive credit cards, who are likely to go into default when given a loan etc. These organisations are not likely to have data to answer questions like what customers think of their products and services, which companies are likely to perform well in the stock market in the near future, and how live updates on critical events can directly impact businesses. The answers to those questions lie scattered all over the web on sites like Lowyat.NET, Reddit, The Star Online, The Edge Markets and so on. Most of the data needed can be obtained by using automated systems to harvest data from selected sites.



THE APPLICATIONS OF WEB SCRAPING/CRAWLING

The applications of web scraping/crawling for banking and financial insitutions are vast.

Competitor price research

To **monitor** the performance of competitors' funds, such as unit trusts

Equity research

To **gain insights** of a few stocks compared against the industry average

Financial ratings

To get **live updates** and drive **high-velocity research** and analytics

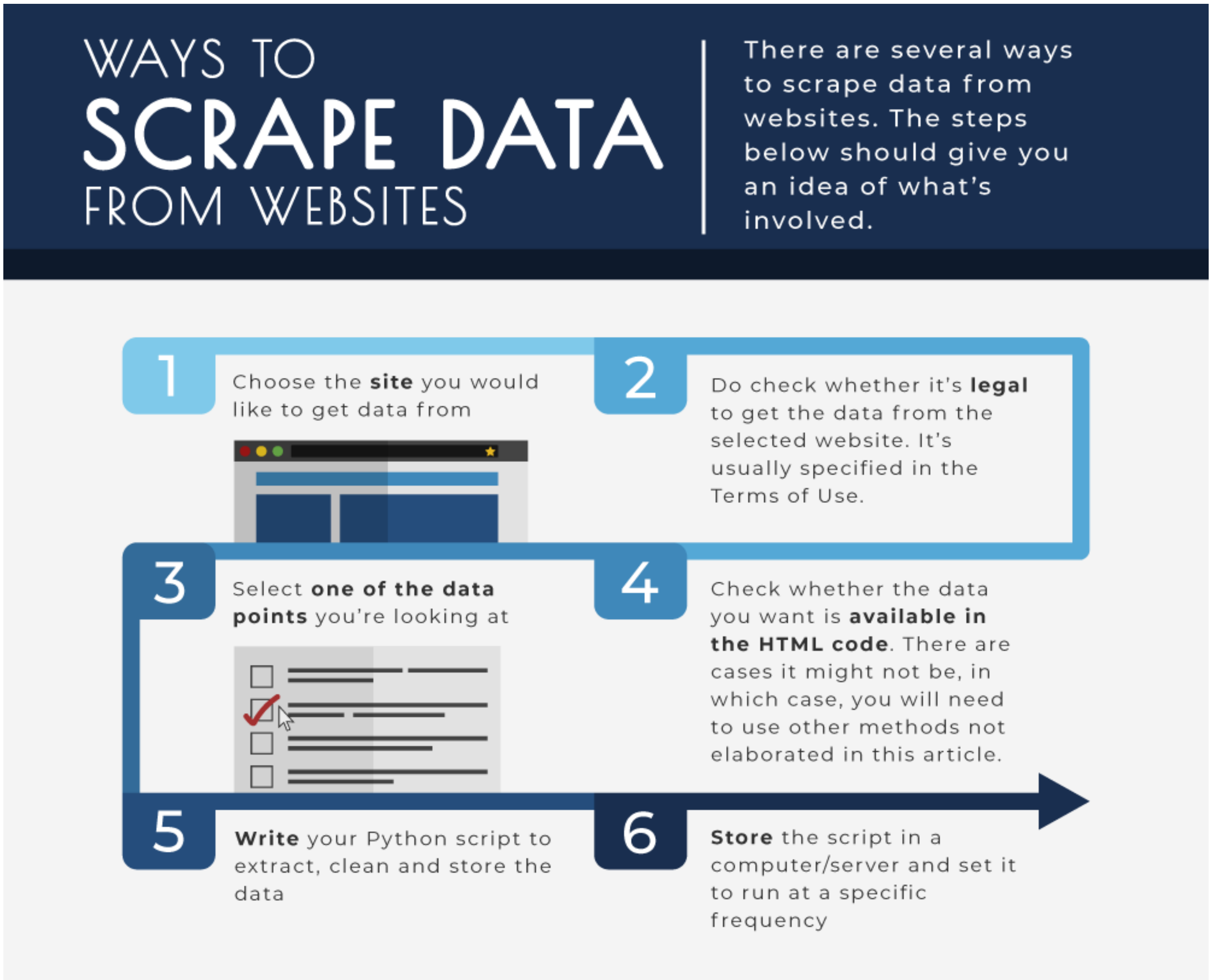
Market sentiment

To **gauge how users feel** about your organisation's products and services from public forums

Web-scraping isn't too useful when you only want to copy and paste information from a website on a one-off basis.

The efficiency gains can only be observed when you have to get that information every 5 minutes in every day or in near real time.

The objective is for you to write the code, store it somewhere, and let it run at a frequency you choose, so you only spend time on the analysis from then on.



WAYS TO SCRAPE DATA FROM WEBSITES

There are several ways to scrape data from websites. The steps below should give you an idea of what's involved.

1 Choose the **site** you would like to get data from

2 Do check whether it's **legal** to get the data from the selected website. It's usually specified in the Terms of Use.

3 Select **one of the data points** you're looking at

4 Check whether the data you want is **available in the HTML code**. There are cases it might not be, in which case, you will need to use other methods not elaborated in this article.

5 **Write** your Python script to extract, clean and store the data

6 **Store** the script in a computer/server and set it to run at a specific frequency

Say if I were to be looking for cheap dividend stocks to invest in, the criteria I would look for are:
1) low or moderate Price-to-Earnings (PE)
2) high Dividend Yield (DE)
3) high Return on Equity (ROE)

Bursa Malaysia's Equities Prices page doesn't have this information, but I did find another site which has on malaysiastock.biz . The ratios I'm looking for are available, but the stocks are listed alphabetically, so I would not be able to see the ratios across the entire stock market. This is where web scraping can help, and the steps below show how to accomplish this.

1. Import the relevant Python libraries

```
import pandas as pd
import requests
import urllib.request
import time
from bs4 import BeautifulSoup ##other options are scrapy and Selenium
```

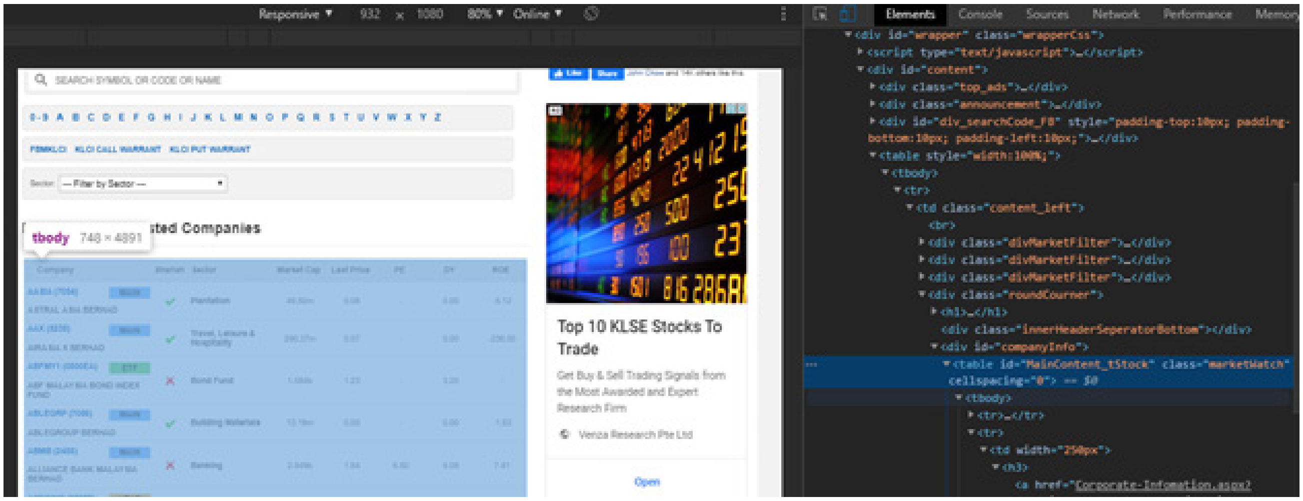
- https://www.bursamalaysia.com/market_information/equities_prices
- https://www.malaysiastock.biz/Listed-Companies.aspx?type=A&value=A

2. Scrape malaysiastock.biz

```
url = ‘https://www.malaysiastock.biz/Listed-Companies.aspx?type=A&value=A’
##scrape only stocks starting with A
response = requests.get(url)
response
<Response [200]> ##this confirms the connection is successful
```

```
soup = BeautifulSoup(response.text, “html.parser”)
##now use BeautifulSoup’s html.parser to display the data
table = soup.findAll(‘table’, {‘class’: ‘marketWatch’})
##and look for a table called marketWatch
table
```

Using Chrome, if you were to right click on **AASIA (7054)**, and click **Inspect**, you will be able to see the html code of the webpage, as snapshot below. Since the data on the website is displayed in a table, it would be reasonable to look for a table within the html code. While parsing the html code, notice the table on the webpage is highlighted when you hover your cursor over **`<table id=“MainContent_tStock” class=“marketWatch” cellspacing=”0”>`**.



```
▾<div id="wrapper" class="wrapperCss">
  ▸<script type="text/javascript">…</script>
  ▾<div id="content">
    ▸<div class="top_ads">…</div>
    ▸<div class="announcement">…</div>
    ▸<div id="div_searchCode_FB" style="padding-top:10px; padding-
    bottom:10px; padding-left:10px;">…</div>
    ▾<table style="width:100%;">
      ▾<tbody>
        ▾<tr>
          ▾<td class="content_left">
              <br>
            ▸<div class="divMarketFilter">…</div>
            ▸<div class="divMarketFilter">…</div>
            ▸<div class="divMarketFilter">…</div>
            ▾<div class="roundCourner">
              ▸<h1>…</h1>
                <div class="innerHeaderSeperatorBottom"></div>
              ▾<div id="companyInfo">
…               ▾<table id="MainContent_tStock" class="marketWatch"
                  cellspacing="0"> == $0
                  ▾<tbody>
                    ▸<tr>…</tr>
                    ▾<tr>
                      ▾<td width="250px">
                        ▾<h3>
                            <a href="Corporate-Infomation.aspx?
```

3. The next part is the pre-processing bit. This is usually deemed the most laborious part of this process. Most data which comes from a html.parser is usually in html format, and it not easily converted into a neat table.

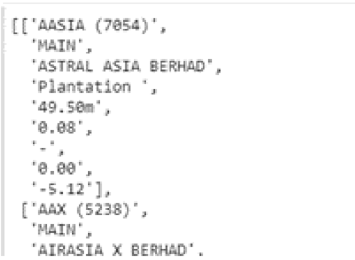
```
example = list(table[0])  ##converts table into a list
example2 = [x for x in example if x != “\n”]  ##get rid of empty lines

prices = []  ##creates an empty array to store data

for i in range(1, len(example2)):
entry = example2[i].get_text(separator=”\n”).split(“\n”)
entry_filtered = [x for x in entry if x]
prices.append(entry_filtered)

prices
```

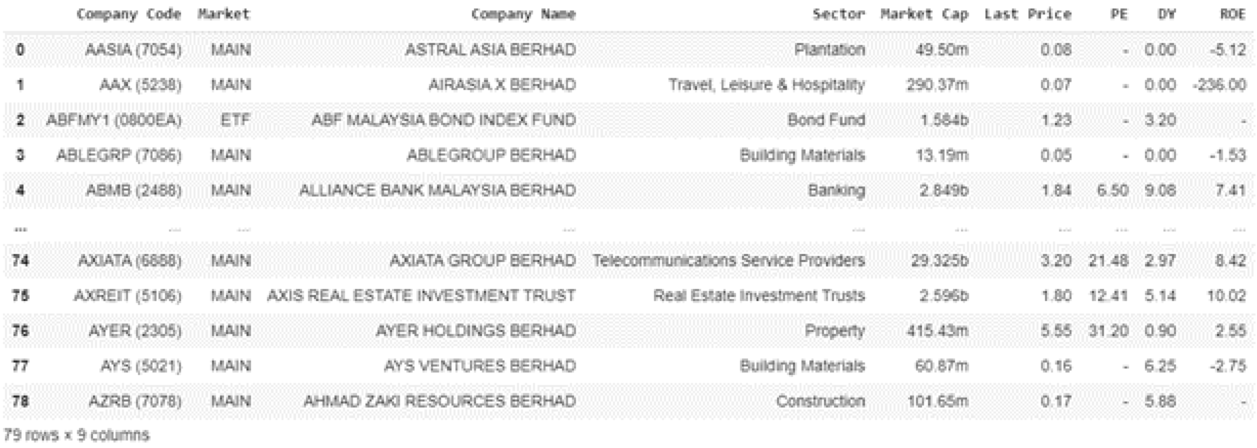
The for loop is for me to separate the data using the delimiter \n , get rid of any more delimiters \n, and to append the result in the prices array. A snapshot of the prices array is to the right.



```
[['AASIA (7054)',
  'MAIN',
  'ASTRAL ASIA BERHAD',
  'Plantation ',
  '49.50m',
  '0.08',
  '-',
  '0.00',
  '-5.12'],
 ['AAX (5238)',
  'MAIN',
  'AIRASIA X BERHAD'.
```

4. Once your data is in an array, it's easy to convert it into a table and to rename the columns. A snapshot of the resulting table is below.

```
complete_table = pd.DataFrame(prices)
complete_table.rename(columns = {0: "Company Code", 1: "Market", 2: "Company
Name", 3: "Sector", 4: "Market Cap", 5: "Last Price",
                                6: "PE", 7: "DY", 8: "ROE"})
```



	Company Code	Market	Company Name	Sector	Market Cap	Last Price	PE	DY	ROE
0	AASIA (7054)	MAIN	ASTRAL ASIA BERHAD	Plantation	49.50m	0.08	-	0.00	-5.12
1	AAX (5238)	MAIN	AIRASIA X BERHAD	Travel, Leisure & Hospitality	290.37m	0.07	-	0.00	-236.00
2	ABFMY1 (0800EA)	ETF	ABF MALAYSIA BOND INDEX FUND	Bond Fund	1.584b	1.23	-	3.20	-
3	ABLEGRP (7086)	MAIN	ABLEGROUP BERHAD	Building Materials	13.19m	0.05	-	0.00	-1.53
4	ABMB (2488)	MAIN	ALLIANCE BANK MALAYSIA BERHAD	Banking	2.849b	1.84	6.50	9.08	7.41
...	...	...	...	...	...	...	...	...	...
74	AXIATA (6888)	MAIN	AXIATA GROUP BERHAD	Telecommunications Service Providers	29.325b	3.20	21.48	2.97	8.42
75	AXREIT (5106)	MAIN	AXIS REAL ESTATE INVESTMENT TRUST	Real Estate Investment Trusts	2.596b	1.80	12.41	5.14	10.02
76	AYER (2305)	MAIN	AYER HOLDINGS BERHAD	Property	415.43m	5.55	31.20	0.90	2.55
77	AYS (5021)	MAIN	AYS VENTURES BERHAD	Building Materials	60.87m	0.16	-	6.25	-2.75
78	AZRB (7078)	MAIN	AHMAD ZAKI RESOURCES BERHAD	Construction	101.65m	0.17	-	5.88	-

79 rows × 9 columns

5. The resulting table only contains the statistics of stocks starting with A. The remaining tasks are to construct a for loop, to run this algorithm for stocks starting with B, C, D, and so on. This can be achieved by running the algorithm above and changing the url variable to: 'https://www.malaysiastock.biz/Listed-Companies.aspx?type=A&value=B' for stocks starting with B, and doing the same for other alphabets. The full code can be accessed via the QR on the right or the link below .

SCAN ME

To deploy this code:

1. Adjust the output directory of the code to point to a data warehouse
2. Store this script in a Docker container, and set the cron job to run it every 15 minutes

Web scraping is useful for other sorts of data too:

1. To gauge what customers think of your organisation's products and services based on comments from public forums like Lowyat.NET and Reddit using sentiment analysis
2. To find out what topics are trending on Twitter and Instagram for product development using topic modelling
3. To gather information from news sites which may impact your organisation's products and services like news coverage about competitor's products

One important takeaway is to identify the problem your business is trying to solve. If it involves web scraping, it's vital to understand the structure of the website on which you would like to scrape the information. Scraping different sites will involve multiple pre-processing methods, which means someone in your team will need to maintain the scraper if the target site's structure changes.

- https://github.com/atlas-github/malaysiastockbiz_scraper/blob/master/malaysiastock_biz_scraper.ipynb

ALTERNATIVE DATA & CREDIT SCORING FOR THE UNBANKED

By Peter Kua Seng Choy

Alternative Data and Credit Scoring for the Unbanked

Credit score, the numerical representation of an individual's creditworthiness, didn't exist until the 1950s. Then, loan officers at banks decided - purely based on their own judgement - whether an individual qualifies for a loan or not. The approval process was biased and subjected to racial discrimination and favouritism.

With the introduction of credit scoring in 1956, the entire loan exercise was standardised. Based on data points such as the individual's payment history, amount of debt, credit age, and credit types, the credit score is calculated. This score helps lenders evaluate the candidate's credit risk and their ability to repay a loan.

Recently, credit scoring saw the introduction of alternative data as an additional category of information. This data further refines the eligibility of a candidate for loans by assessing factors outside the traditional credit scoring datasets. Incorporating alternative data into credit scoring can increase the overall accuracy in evaluating a person's financial standing.

What is Alternative Data in Credit Scoring?

Alternative data is information collected from non-traditional sources that helps financial service providers gain a complete view of an individual's creditworthiness. While traditional credit data includes an individual's credit history and debt amount, alternative credit data includes information such as rental and utility payment history, asset ownership, alternative financial data, and shopping history.

Why is Alternative Data Important in Credit Scoring?

Alternative credit data helps lenders expand their services to "credit-invisibles"; people who were previously unqualified for loans based on the conventional credit scoring system. As traditional credit data reports uphold a person's credit history as a decisive factor for scoring, an individual without credit history will have a tough time qualifying for new credit. With alternative data, credit-invisible consumers have improved chances of obtaining a loan.

Alternative Data Can Increase Credit Scoring Accuracy

A primary concern for lenders is the risks associated with each candidate. Even if an individual has a perfect credit score, there is a chance that he or she poses a threat. For example, if an individual maintains a decent credit score but defaults on insurance payments, this person is considered high risk. Yet traditional credit data fails to assess this.

Taking advantage of alternative data, lenders can gain crucial insights into candidates even if they are eligible under conventional credit scores. Lending decisions become more precise.

For consumers who are already eligible for loans under the traditional credit scoring system, alternative data can further improve their credit scores, qualifying them for more attractive interest rates.

Consumers cannot easily influence conventional credit data. With alternative credit data, however, an individual is empowered to contribute their rental or insurance payment history to shape their credit score.

Having said that, financial institutions would do well by combining both traditional and alternative credit data sources to create more accurate credit risk models. They would then be able to better predict the risk of an individual or business defaulting on a loan.

Examples of Alternative Data That Can be Used

To incorporate alternative data into credit scoring, it must be accessible for analysis. The data must also be a good predictor of credit behaviour and comply with all laws associated with consumer credit evaluation.

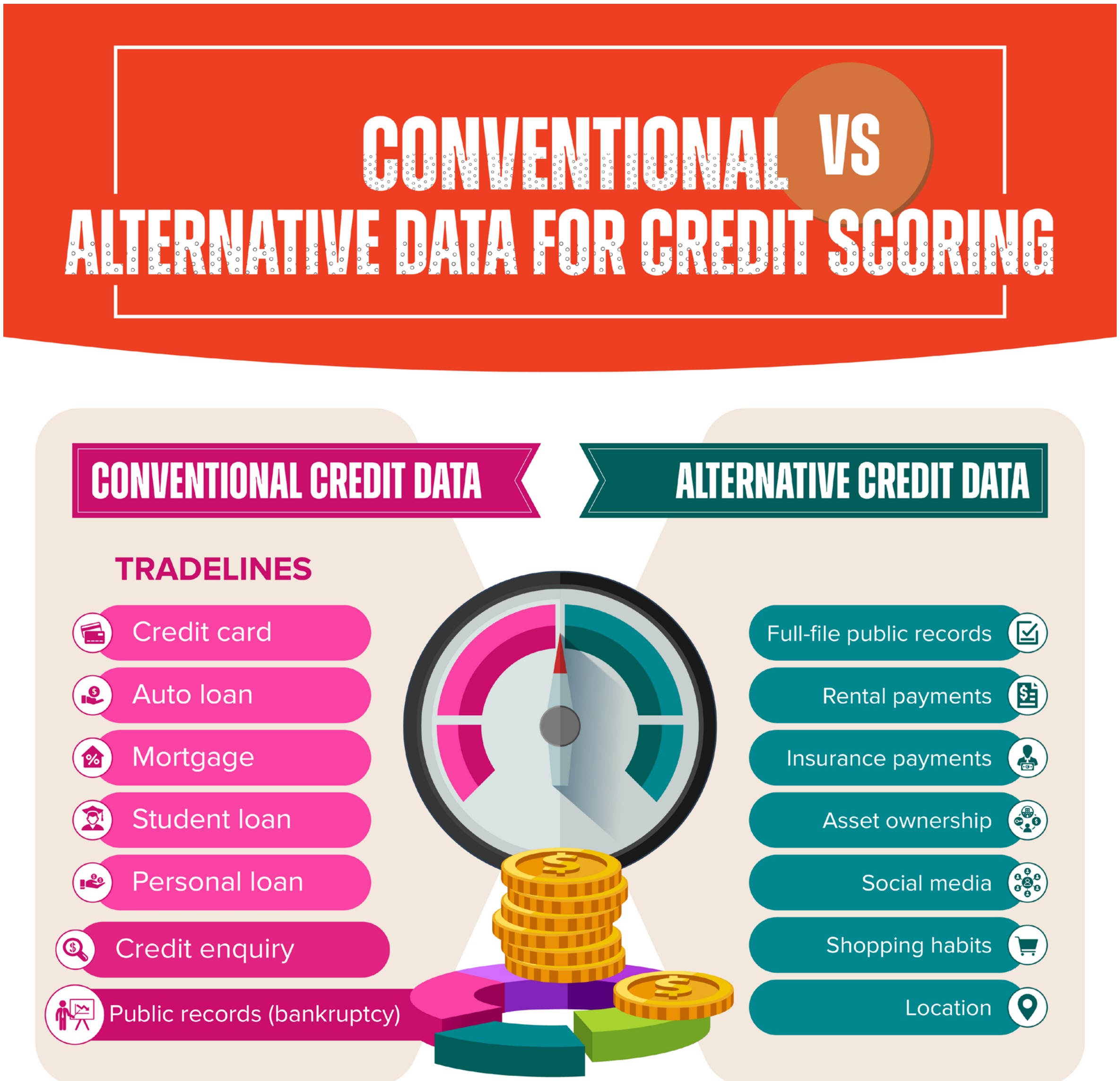
Data analytics company FICO uses a six-point-test to determine whether any new form of data is worthy of inclusion into the credit scoring system. The test covers the following key dimensions:

- Regulatory Compliance - Data must comply with all regulations associated with consumer credit evaluation.
- Depth of Information - The more in-depth and broader a set of data, the higher its consideration will be.
- Accuracy - Data collected must be accurate; otherwise, it compromises predictiveness.
- Predictiveness - Data should be capable of predicting a consumer's future repayment behaviour.
- Consistency - Data must be consistent and did not undergo significant changes.
- Additive Value - Data must supplement or complement the information already used in credit bureau reports.

Here are some examples of alternative data for credit scoring:

- Full-file public records of an individual.
- Utility bill payment history for services including water, electricity and gas.
- Rental payment history.
- Insurance payment history.
- Information on alternative financial services used such as micro loans, point-of-sale financing and title loans.
- Financial account aggregation, which contains combined information from different books, such as bank and investment accounts.

For commercial loans, a business' location, amenities and accessibility can be used as alternative credit data. Lenders can also consider social media and online information. If a retailer has excellent ratings online and social networking platforms wax lyrical about their products, then they are likely to have good ROI. Such a business should be reliable in terms of loan repayments.



CONVENTIONAL VS ALTERNATIVE DATA FOR CREDIT SCORING

CONVENTIONAL CREDIT DATA

TRADELINES

- Credit card
- Auto loan
- Mortgage
- Student loan
- Personal loan
- Credit enquiry
- Public records (bankruptcy)

ALTERNATIVE CREDIT DATA

- Full-file public records
- Rental payments
- Insurance payments
- Asset ownership
- Social media
- Shopping habits
- Location

Final Thoughts

Employing alternative credit data to generate credit scores has remarkable benefits for both consumers and lenders.

Alternative credit data offers consumers with scant credit histories hope for obtaining loans. This data also allows people to add more value to their existing credit scores by including additional financial information not covered by traditional credit data.

For financial institutions, alternative credit data brings more accuracy to the lending system and can help them better understand the repayment capacity of an individual or business.

For training enquiries, please contact:

Asian Banking School (201201039737 (1024215-T))
Level 3, Bangunan AICB
10, Jalan Dato Onn
50480 Kuala Lumpur, Malaysia

Tel : +603-2742 7822/21 **E-mail :** digitalbanking@asianbankingschool.com

Asian Banking School

Asian Banking School (ABS)

www.asianbankingschool.com