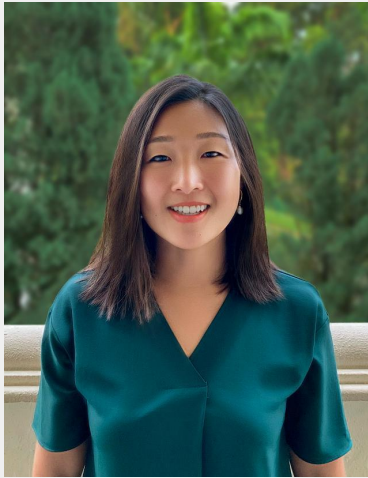




Merchant Master: an Efficient Data Extractor & Mining Tool

The Team



Daphne Teh

- Management PhD
- Experience working with large datasets



Ronja Harder

- MA International Relations
- Experience in foreign policy and security



Ziba Mirza

- MBA and MSc in Development
- Experience in research and HR

Garbage In Garbage Out



=



The Challenge



...is a major Swiss financial service provider

Actions

1. **Extracting** cleaned merchant names
2. **Mining** additional information about merchants
3. **Categorizing** merchants

Outcomes

Produce High Quality Big Data
Derive Insights

AIR SERBI0123456789

7A Factory

*DEOVGROUPLT

*ANTOINEMED18

TEA ROOM L'AQUARIUM

Extracting Cleaned Merchant Names



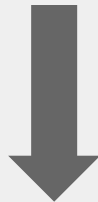
Extract

Mine

Categorize

Outcomes

Original # rows: 2418



Clean # rows: 2217

Merchant Name

SIXT 0123456789/M

*S&I MARINCAMP

*4149Code

merchant

processor

```
def merchant_name(df, column_name):  
    df['id'] = df[column_name].str.lower()  
    df = add_payment_gateway(df, column_name, payment_gateway)  
    df = add_legal_form(df, column_name, legal_form)  
    df = add_website_column(df, column_name)  
    df = remove_numbers(df, column_name)  
    df = remove_special_characters(df, column_name)
```

Clean Merchant Name

Sixt

S I Marincamp

~~Code~~

Mining Metadata about Merchants



Extract

Mine

Categorize

Merchant name
Doldenhornhutte
Minimum Ag



Google

```
def mine_google(merchant_names):  
    df = pd.DataFrame(columns=['merchant'  
  
    for merchant_name in merchant_names:
```

● 1,214 hits (50%)
■ 33 - 54% return on hits



Open Street Maps

```
def osm_by_address(addresses):  
    df_original = pd.read_csv('clear  
  
    df_locaddress = pd.DataFrame()
```

● 839 hits (38%)
■ 95% return on hits



Zefix (Swiss Govt)

```
def mine_zefix(df_merchant):  
    merchant_names = df_merchant[  
  
    result = pd.DataFrame()
```

● 1240 hits (56%)
■ 27 - 55% return on hits

Merchant name	Reviews	Address	Opening hours	Business type	Canton	Latitude	Longitude	Address	Legal form	CH-ID
Doldenhornhutte	4.8	3718 Kandersteg		Mountain cabin	Bern/Berne	46.497587	7.675664	Biberg 5	General Partnership	CH-036-2077631-1
Minimum Ag	4.4	Flüelastrasse 31	{Monday: 8am-11pm}		Zürich	47.381651	8.495320	Flüelastrasse 31	Corporation	CH-020-3037161-4

Categorizing the Merchants: success rate of 66%



Extract

Scrape

Categorize

APIs

● 735 hits (33.2%)

Prioritize words from names

['garage', 'w', 'brumann', 'ag']



Language Detection

['garage'] → English

['blumen'] → German



Translation

['garage'] → 'garage'

['blumen'] → 'flower'



Types of words

['flower'] → noun

['stella'] → null

Curated Lists

● 1057 hits (47.7%)

Salt Center Shop Fribourg



Salt.

Transport

[sbb, öbb, sixt, vbz, train]

Beauty

[beauty, belle, coiffure, hair]

Google Results

● 514 hits (34.2%)

Garage W. Brumann Ag

Car dealer → Car Related

Coiffure Kopfform

Beauty salon → Beauty

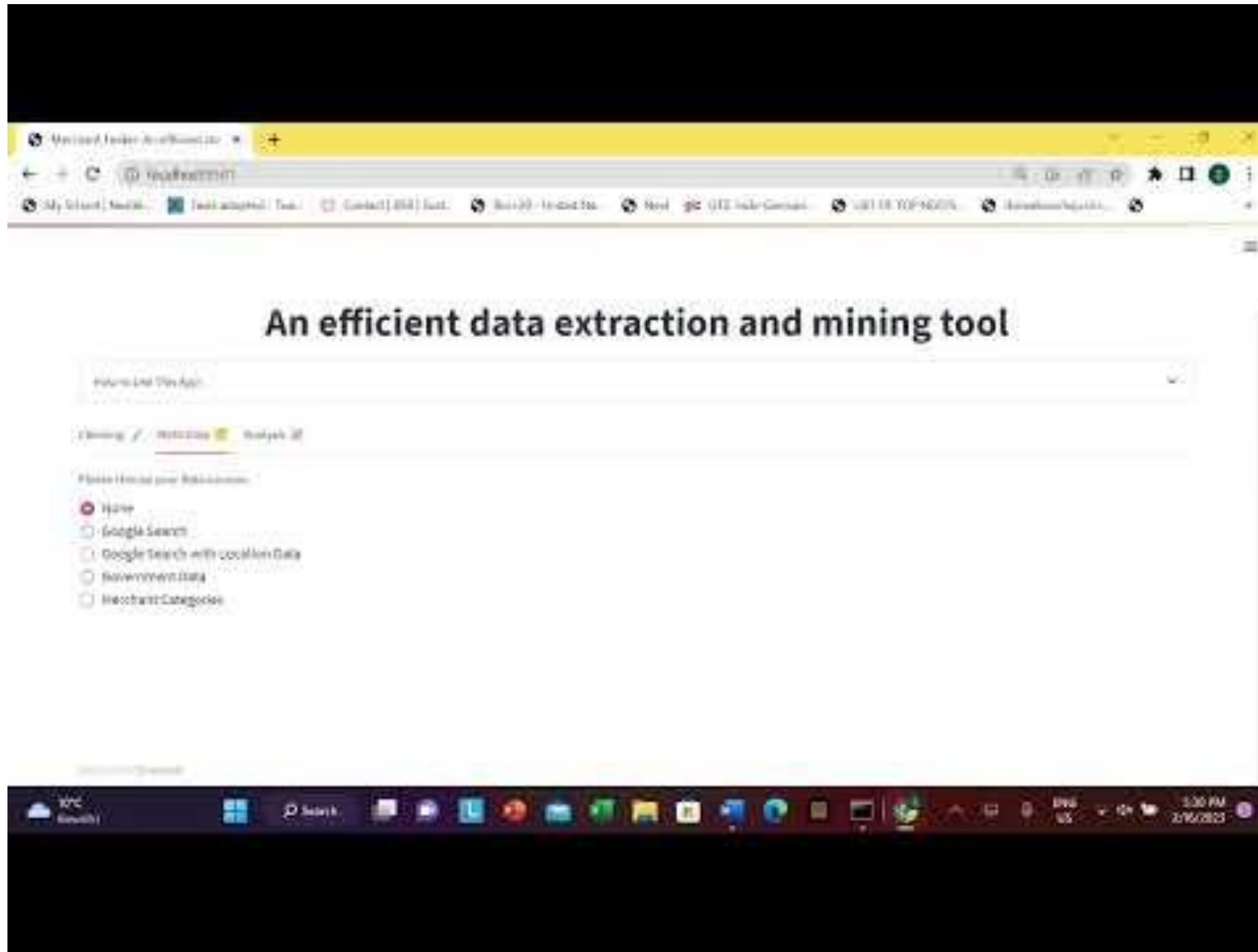
Queen's Beauty Marisa

Cosmetic **store** → Shop

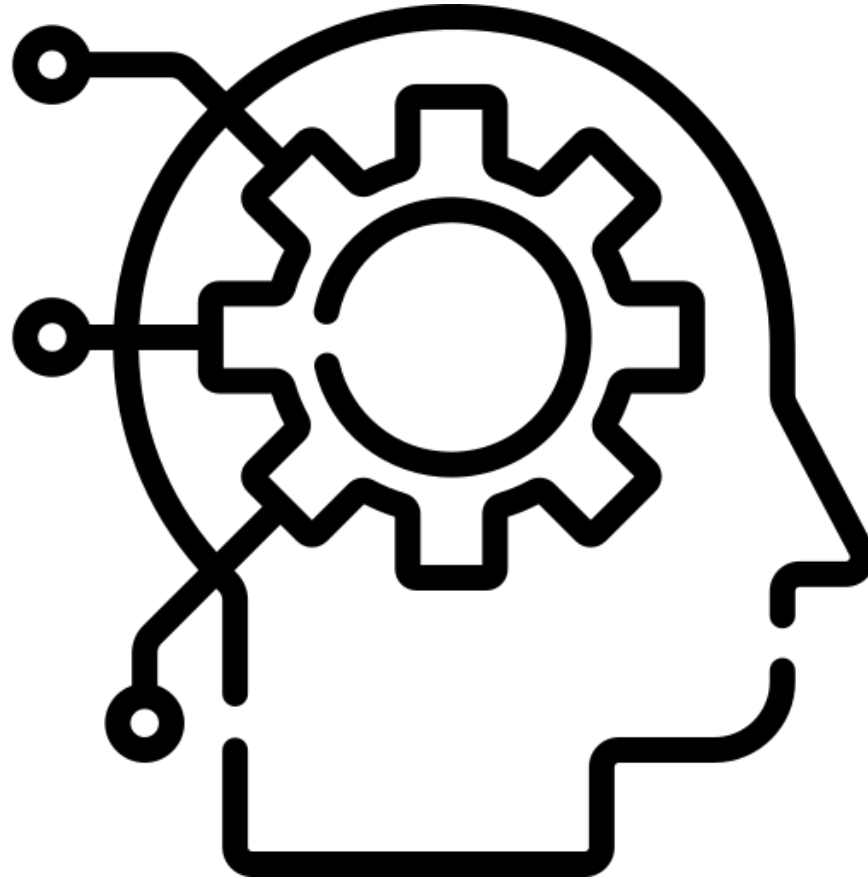
Activ Fitness Ag Affoltern

Fitness Center → Sport

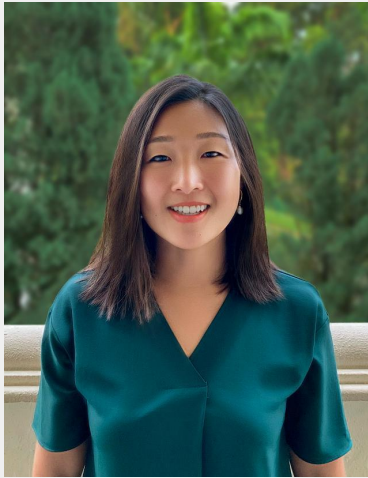
Integrating the Solution: Our Merchant Master



Key Take-aways and Further Steps



How to Contact Us



Daphne Teh

daphne.s.teh@gmail.com

<https://www.linkedin.com/in/daphne-s-teh/>



Ronja Harder

ronjaharder@gmail.com

<https://www.linkedin.com/in/ronjaharder/>



Ziba Mirza

zibamirza@gmail.com

<https://www.linkedin.com/in/ziba-mirza/>



Questions

and thank you!