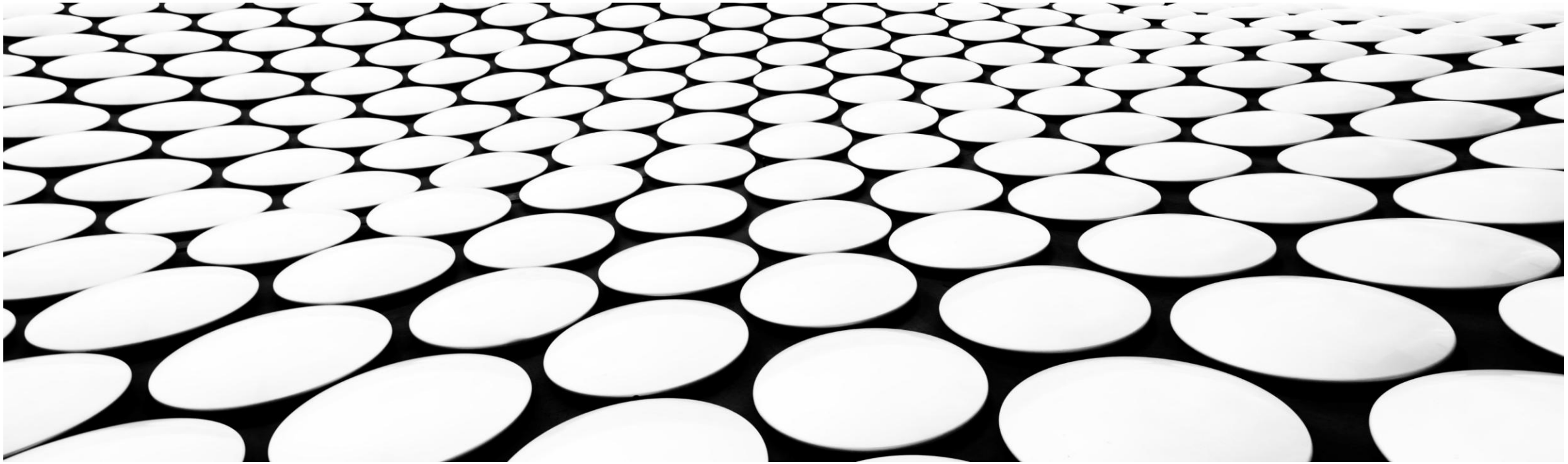
# DATA COLLECTION

*Comparing approaches*



# DATA COLLECTION

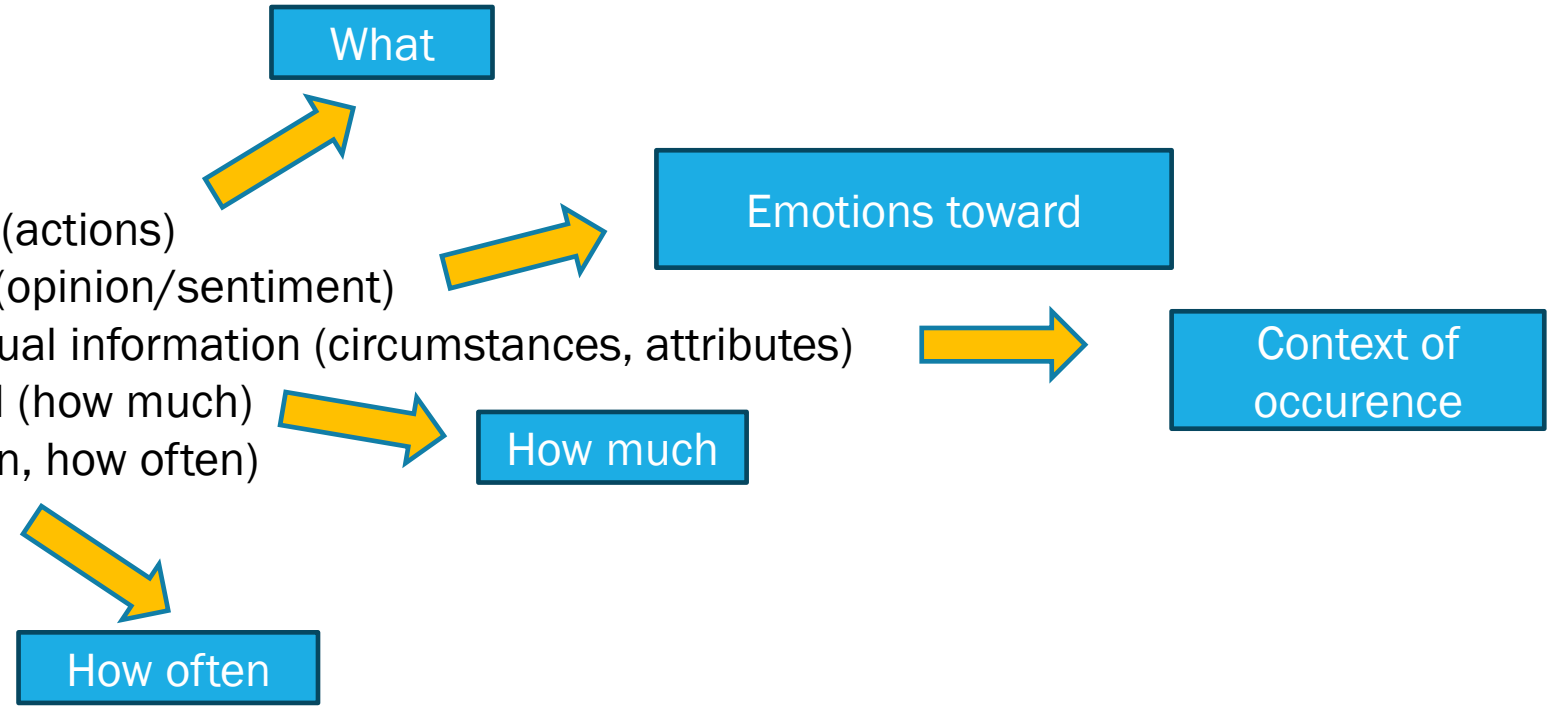
- Decision Factors for Data Collection Method Selection
- Three comparative case studies

# DATA COLLECTION

# What are we trying to capture?

People

1. Behavioral information (actions)
2. Attitudinal information (opinion/sentiment)
3. Environmental/Contextual information (circumstances, attributes)
4. Financial/Transactional (how much)
5. Frequency/Usage (when, how often)



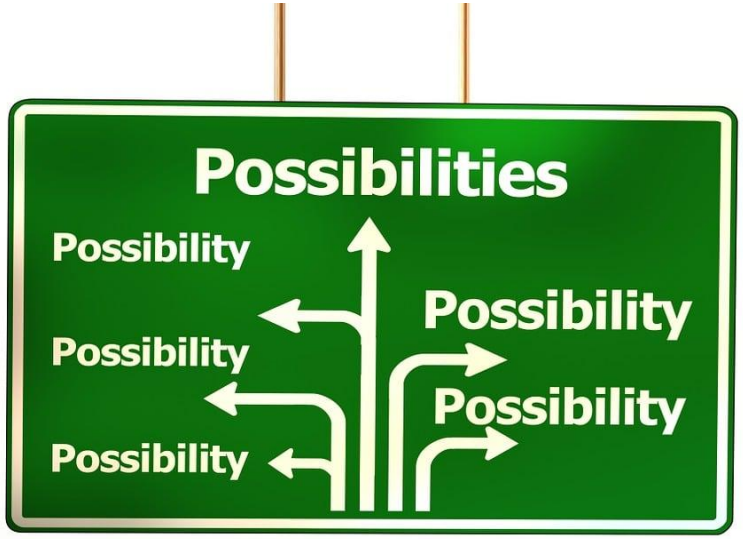
| Self-Reported Data | Surveys<br>Interviews<br>Focus groups |
| --- | --- |
| Data directly provided by individuals | |

| Behavioral and Observational Data | Observations<br>Experiments |
| --- | --- |
| Data collected through monitoring or observing human behavior. | |

| Automated Data | Logs<br>Sensor<br>Transactions |
| --- | --- |
| Data generated obtained through an automated system. | |

| Digital and Platform-Based Data | Social Media Data<br>Web Scraping<br>APIs |
| --- | --- |
| Data originating from digital platforms, systems, or online activity. | |

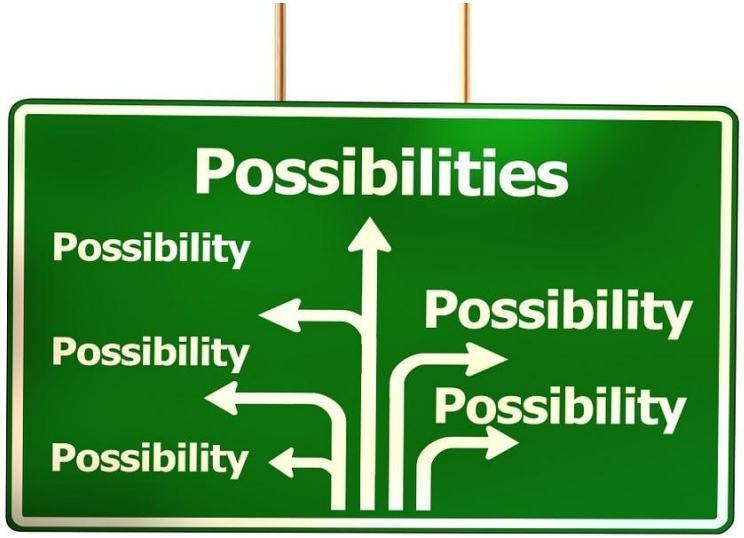
# DECISION FACTORS




Quite overwhelming… How to choose/decide on data collection method?

- Resource constraints (Effort/Cost)
- Time constraints
- Goal of the study
- Availability of already collected data
- Domain expertise
- Technical expertise
- Data storage possibilities
- Ease of use and integration
- Reliability
- Ethical and Privacy concerns

## Data Collection




1. Purpose (Problem Statement/Question/Goal of Study)
2. Resources (Time/Effort/Cost)
3. Availability of already collected data
4. Domain Expertise
5. Technical Expertise
6. Data Storage possibilities (cost)
7. Reliability
8. Ease of use (feasibility) and integration
9. Ethical and Privacy concerns

# 1. Purpose

*Which collection method would allow us to answer our research question?*

Easy

What is the expected bounce rate if the company adopts the suggested website redesign?

A/B Testing

Not so easy

How much money are younger people spending on clothing compared to older people?

Survey
Observation
Logs
Transactions
Social Media

## 2. Resources

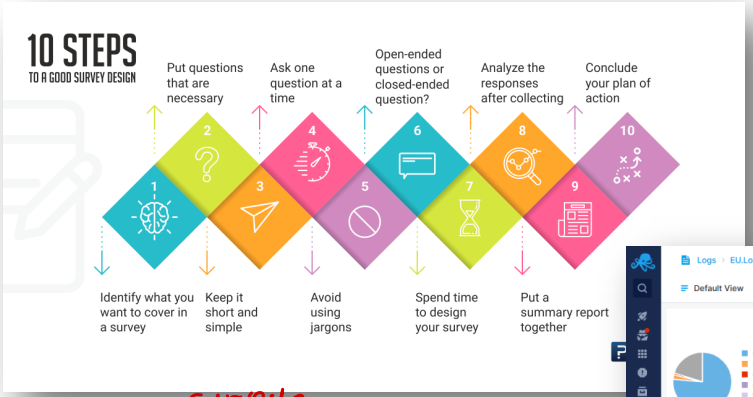
*How much money, time, equipment do we have?*

<u>Not so easy</u>

How much money are younger people spending on clothing compared to older people?

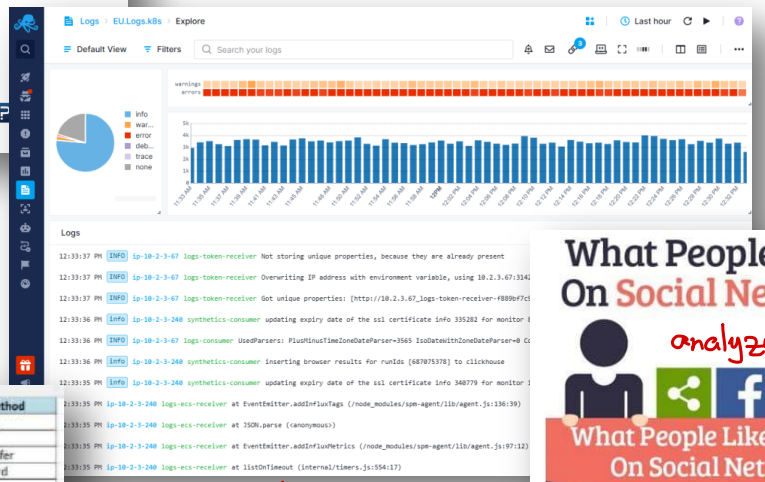
Survey
Observation
Logs
Transactions
Social Media

Survey
Observation
Logs
Transactions
Social Media




surveys

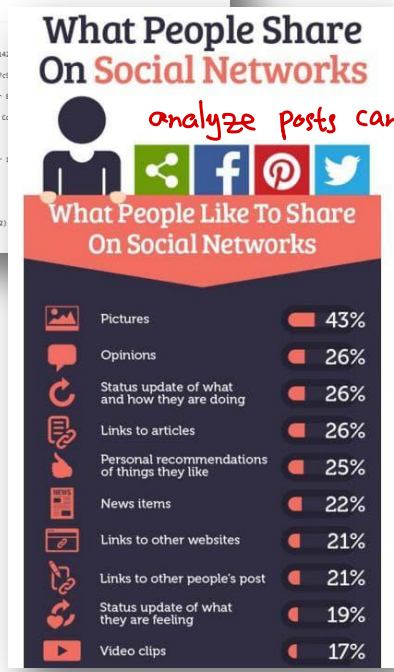

observational




logs

| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| 1 | Date | Transaction ID | Customer Name | Category | Amount | Payment Method |
| 2 | 2023-01-01 00:00:00 | T00001 | John Doe | Forex | 54697.34 | Check |
| 3 | 2023-01-01 11:40:32 | T00002 | Alice Brown | Equity | -6230.63 | Cash |
| 4 | 2023-01-01 23:21:05 | T00003 | Jane Smith | Derivative | 99837.92 | Wire Transfer |
| 5 | 2023-01-02 11:01:37 | T00004 | Jane Smith | Commodity | 4473.36 | Credit Card |
| 6 | 2023-01-02 22:42:10 | T00005 | Jane Smith | Fixed Income | 71438.19 | Cash |
| 7 | 2023-01-03 10:22:42 | T00006 | Bob Johnson | Forex | 80311.67 | Credit Card |
| 8 | 2023-01-03 22:03:15 | T00007 | Jane Smith | Forex | 31036 | Check |
| 9 | 2023-01-04 09:43:47 | T00008 | Charlie Green | Forex | 11653.73 | Wire Transfer |
| 10 | 2023-01-04 21:24:19 | T00009 | John Doe | Derivative | 868.59 | Wire Transfer |
| 11 | 2023-01-05 09:04:52 | T00010 | Alice Brown | Commodity | 72346.66 | Credit Card |
| 12 | 2023-01-05 20:45:24 | T00011 | Alice Brown | Forex | 39791.89 | Credit Card |
| 13 | 2023-01-06 08:25:57 | T00012 | John Doe | Commodity | 68508.95 | Wire Transfer |
| 14 | 2023-01-06 20:06:29 | T00013 | Charlie Green | Fixed Income | 90694.84 | Credit Card |
| 15 | 2023-01-07 07:47:02 | T00014 | Bob Johnson | Equity | 6124.21 | Wire Transfer |
| 16 | 2023-01-07 19:27:34 | T00015 | Alice Brown | Equity | 91108.81 | Online Banking |
| 17 | 2023-01-08 07:08:06 | T00016 | Bob Johnson | Fixed Income | 35278.91 | Check |
| 18 | 2023-01-08 18:48:39 | T00017 | Bob Johnson | Equity | 23579.37 | Wire Transfer |
| 19 | 2023-01-09 06:29:11 | T00018 | Alice Brown | Forex | 93736.85 | Check |
| 20 | 2023-01-09 18:09:44 | T00019 | Bob Johnson | Derivative | 98971.69 | Check |
| 21 | 2023-01-10 05:50:16 | T00020 | John Doe | Equity | 11878.14 | Online Banking |
| 22 | 2023-01-10 17:30:49 | T00021 | Jane Smith | Fixed Income | 62252.22 | Wire Transfer |
| 23 | 2023-01-11 05:11:21 | T00022 | Alice Brown | Derivative | 1714.48 | Credit Card |
| 24 | 2023-01-11 16:51:54 | T00023 | Jane Smith | Fixed Income | 61600.54 | Credit Card |
| 25 | 2023-01-12 04:32:26 | T00024 | Alice Brown | Forex | 81004.46 | Wire Transfer |
| 26 | 2023-01-12 16:12:58 | T00025 | Charlie Green | Fixed Income | 65294.84 | Wire Transfer |
| 27 | 2023-01-13 03:53:31 | T00026 | Jane Smith | Forex | 35906.65 | Check |
| 28 | 2023-01-13 15:34:03 | T00027 | Alice Brown | Equity | 32137.3 | Check |
| 29 | 2023-01-14 03:14:36 | T00028 | Bob Johnson | Derivative | 33243.47 | Cash |
| 30 | 2023-01-14 14:55:08 | T00029 | Jane Smith | Commodity | 54868.3 | Check |

transactions




analyze posts can be hard

social media
requires technical expertise

## 3. Availability of already collected data

*Is there already available data to help us answer our research question?*

we need to evaluate the research question

**Types of Data Sources: A Comprehensive Guide to Understanding Different Data Sources**

Sumana Dotnettricks · Follow
10 min read · Jun 13, 2023

In this article, they also highlight secondary data

Source

| Secondary or Pre-existing Data | Government Databases<br>Academic Research<br>Industry Reports<br>Publicly Available Data<br>Media Sources |
| --- | --- |
| Pre-existing data collected by someone else for a different purpose | |

Statistics Canada

Government Database

The United States Census Bureau's database offers comprehensive demographic data, allowing researchers to examine population trends and patterns across different regions and time periods.

Publicly Available Data

The World Bank's Open Data initiative provides free access to a vast repository of global economic indicators, enabling researchers and policymakers to track and analyze economic trends.

**Replace the collection method**

Is there secondary data that contains **sufficient information** to answer our research question?

**Replace part of the collection method**

Is there secondary data that contains some information that would contribute to answer our research question?

**Provide additional context for study**

Is there secondary data that complements the collection data (e.g. providing additional context) ?

Example of Loyalty program of clothing store. When signing customers might have included their age and/or their postal code. Different aspects of demographics can then be inferred from postal code, as published by some government (e.g. provincial, municipal) data.

# 4. Domain Expertise

*Do we (anyone in our team) know about the area of study?*

Environment: Air Quality

How do air quality levels vary across different regions of Ottawa during peak traffic hours?

→ Behavioral (non human!) Which sensors to use?

Transportation Services

What are the most common complaints about public transportation services in Gatineau?

→ Opinion (very human!) Which questions are targetting the issues?

# 5. Technical Expertise

*Do we (anyone in our team) have technical expertise in line with the collection method?*

What is the expected bounce rate if the company adopts the suggested website redesign?

→ Can you design and put in place A/B Testing?

How do social media users perceive the new health policy announced by the government?

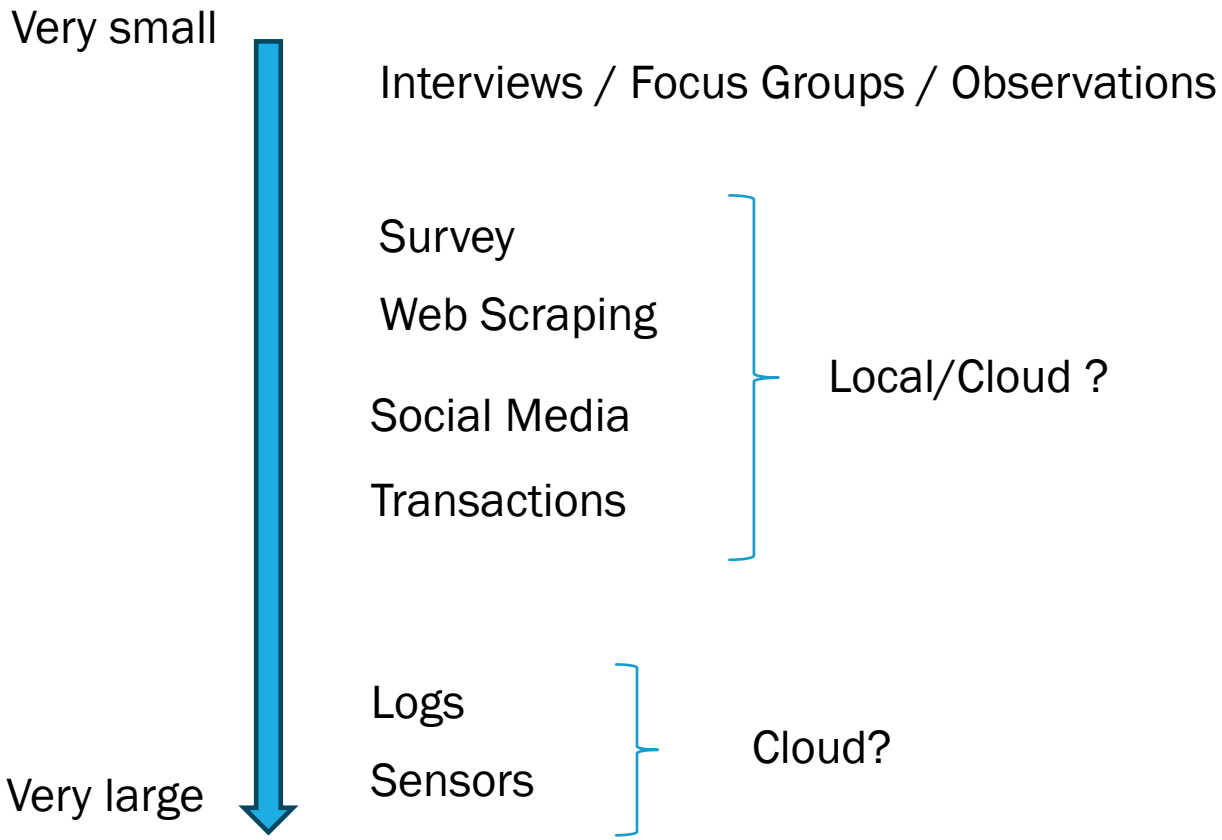
→ Can you design and put in place Information Extraction from text found in Social Media?

What are the primary motivations of employees for choosing remote work over on-site work in a company?

→ Can you design and put in place one-on-one interviews or focus groups to obtain self-reported data?

# 6. Data Storage possibilities

*What is our budget (and current equipment) for local and/or cloud storage?*



# 7. Reliability

*Do we have confidence in the implementation of the collection method and the obtained answers?*

Think of one or two potential issues of each approach?

| | Data Collection | Possible reliability issue |
|---|---|---|
| Self-Reported Data | Survey | Poorly written questions. Users providing impressions based on memory, not necessarily factual. |
| | Interviews | Interviewer not knowledgeable enough. Participants not cooperative. |
| | Focus Groups | Moderator not experienced. Participants not expressing their real opinion. |
| Observational and Behavioral | Observation | Observation is too intrusive. No participant or reluctant participants. |
| | Experiment (A/B) | A/B setup is incorrect (e.g. random not really random). Controlled variable not well identified. |

| | Data Collection | Possible reliability issue |
|---|---|---|
| Automated | Logs | Logging platform not functioning well. Turnover/buffer (amount of data kept) is not sufficient. |
| | Sensors | Sensors malfunctioning. |
| | Transactions | DB not set up properly. |
| Digital and Platform based | Social Media | Extraction method is imperfect. |
| | Web Scraping | Extraction method is not adapted to new version of web site. |
| | APIs | Data provided not from a trustable source. |

## 8. Implementation and integration

*Is the collection method « implementable » (feasible) to answer the research question? If so, will the data collected be easily usable?*

Question: What are the primary motivations of employees for choosing remote work over on-site work in a company?

Implementation --- Feasibility of the actual collection method

→ Sensors are just not possible

Integration – Once it's collected, can we use it?

→ What to do with interviews conducted? What can be said?

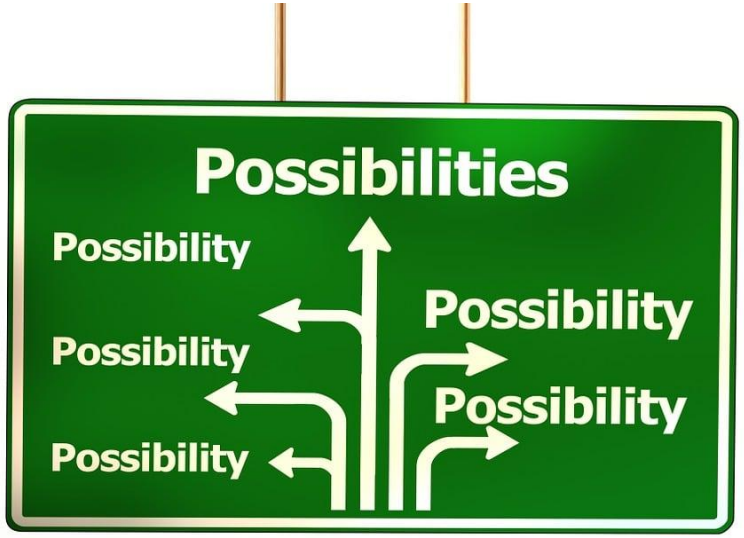
# 9. Ethical and Privacy concerns

*Will the collection method infringe on people's privacy and create ethical concerns?*

How much money are younger people spending on clothing compared to older people?

| Data Collection Method | Example | Ethical concern |
|---|---|---|
| Survey | I provide my age and how much I spend per month on clothing. | I do it from my own will. |
| Observations | I'm being observed in a clothing store. | Did I give my permission for that? |
| Logs | I'm being « followed » through my online searches for various items to buy. | Passive observation (trickier… cookies everywhere). E.g. Google knows where I live and what I click on. |
| Transactions | I buy clothing with my credit card. | My credit card transactions would be linked to my age/clothing spending. Have I agreed to the use of aggregated data? |
| Social Media | I am a follower of a store to know their new collections and sometimes comment on them. | It's a bit like web scraping… legal (shared content), but what about « friends » or « followers »?) |

## Data Collection




1. Purpose (Problem Statement/Question/Goal of Study)
2. Resources (Time/Effort/Cost)
3. Availability of already collected data
4. Domain Expertise
5. Technical Expertise
6. Data Storage possibilities (cost)
7. Reliability
8. Ease of use (feasibility) and integration
9. Ethical and Privacy concerns

# Case Studies

Case Study
Commuting Habits

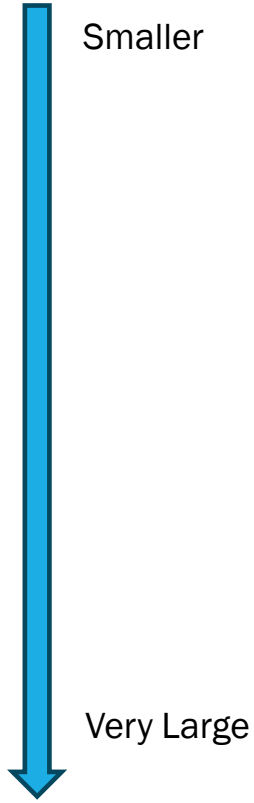
**Goal:** Understand how people commute and why they choose certain transportation modes.

Case Study
Eating Habits

**Goal:** Understand both *what people eat* and *why they eat that way*.

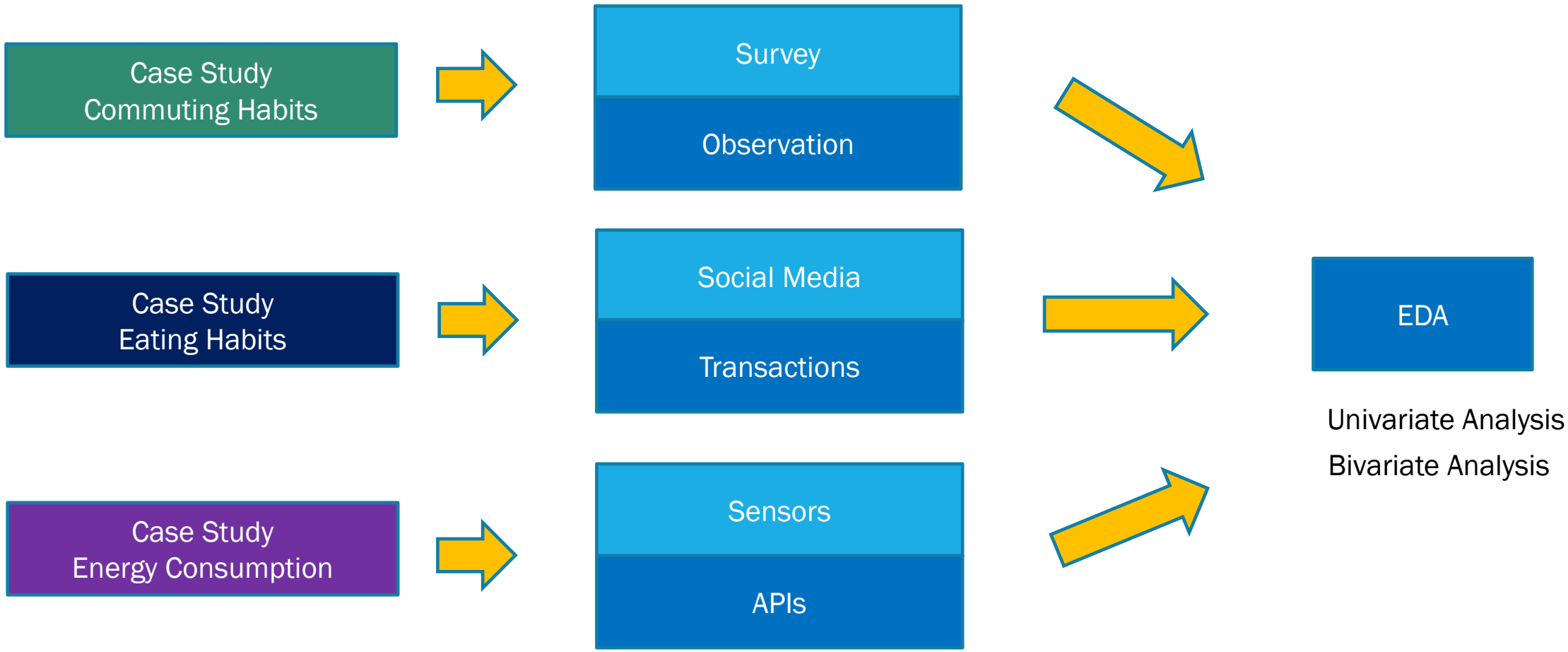
Case Study
Energy Consumption

**Goal:** Understand *energy use behavior*, *perceptions*, and *influencing factors*.



Case studies:

Focus on two collection methods.



# Case Study
# *Commuting Habits*

Survey

Observation

Case Study
Commuting habits

# Survey

Impact on EDA – Numerical/Categorical

## Section 2: Commuting Patterns

4. **What is your primary mode of commuting?**
   - ☐ Car
   - ☐ Bus
   - ☐ Train / Subway
   - ☐ Bicycle
   - ☐ Walking
   - ☐ Other: _____
5. **How long is your average one-way commute?**
   - ☐ Less than 15 minutes
   - ☐ 15–30 minutes
   - ☐ 31–45 minutes
   - ☐ 46–60 minutes
   - ☐ More than 60 minutes

Often request for demographics

## Section 1: Demographics

1. **Age:** _____
2. **Gender:**
   - ☐ Male
   - ☐ Female
   - ☐ Non-binary / Third gender
   - ☐ Prefer not to say
3. **Employment/Study Status:**
   - ☐ Full-time employee
   - ☐ Part-time employee
   - ☐ Student
   - ☐ Other: _____

## Section 3: Preferences & Decision Factors

8. **Which factor is most important when choosing your commuting mode?** *(Select up to 2)*
   - ☐ Cost
   - ☐ Travel time
   - ☐ Convenience
   - ☐ Environmental impact
   - ☐ Comfort
   - ☐ Safety
   - ☐ Other: _____
9. **Do you consider environmental impact when choosing your commuting mode?**
   - ☐ Always
   - ☐ Sometimes
   - ☐ Rarely
   - ☐ Never
10. **Would you be willing to change your commuting mode if it saved money, time, or was more environmentally friendly?**
    - ☐ Yes, definitely
    - ☐ Maybe
    - ☐ No

Does there seem to be a bias here?

## Section 4: Behavior & Technology

**11.** Do you use apps or services to plan your commute?

- ☐ Yes
- ☐ No

**12.** If yes, which apps or services do you use? (Open-ended)

- 

**13.** Do you track your commute time or expenses?

- ☐ Yes, regularly
- ☐ Occasionally
- ☐ No

## Optional Section 5: Open Feedback

**14.** Please share any comments on how commuting could be improved in your area:

- 

Open-ended questions.
Text Analysis.

Observations

Case Study
Commuting habits

### Observational Setting

- Location: Downtown intersection near transit hub
- Time window: 7:30–9:00 AM (weekday)
- Method: Manual counts every 15 minutes

| Date | Time Block | Cars | Buses | Bikes | Pedestrians |
|---|---|---|---|---|---|
| 2025-03-04 | 7:30–7:45 | 128 | 6 | 14 | 52 |
| 2025-03-04 | 7:45–8:00 | 162 | 8 | 21 | 68 |
| 2025-03-04 | 8:00–8:15 | 174 | 9 | 25 | 74 |
| 2025-03-04 | 8:15–8:30 | 158 | 7 | 19 | 63 |
| 2025-03-04 | 8:30–8:45 | 141 | 6 | 17 | 58 |

## Comparison / Integration

| | Survey | Observation |
|---|---|---|
| What is measured? | Self-reported behavior and perceptions | Directly observed commuting behavior |
| Whos is represented? | Survey respondents | People passing through a specific location |
| Level of details | Individual-level | Aggregate |
| Temporal coverage | Typical or average behavior (weekly) | Specific dates and time windows |
| Spatial coverage | Broad (city) | Narrow (intersection) |

# Case Study
## *Eating Habits*

Social Media

Transactions

# Social Media

## Case Study
## Eating Habits

**Post 1 (X / Twitter)**

> "Groceries are wild lately. $70 and somehow still nothing to cook 😩"

**Post 2 (Instagram story caption)**

> "Made a salad for lunch today... then ordered fries at 4pm 🤷‍♀️"

**Post 3 (Reddit – r/EatCheapAndHealthy)**

> "I want to eat better but by the time I get home I just don't have the energy to cook. Anyone else?"

**Post 4 (TikTok comment)**

> "All these 'healthy meals' videos use ingredients I can't even find where I live."

**Post 5 (Facebook neighborhood group)**

> "Any decent takeout places around here that aren't crazy expensive?"

**Post 6 (X / Twitter)**

> "I swear meal prepping only works if you don't mind eating the same thing 5 days in a row."

Posts could provide more info that can be structured

But… the text is tricky to structure

| Target Field | Example |
|---|---|
| PostID | SM01 |
| Platform | X, Instagram, TikTok, Reddit |
| Timestamp | 2025-02-25 |
| Text | "Groceries are ridiculous lately. $80 and I still need dinner ideas." |
| Likes | 124 |
| Shares / Retweets | 18 |
| Hashtags | #groceries, #inflation |

## Regular Expressions

- Extracting anything « regular » in its surface form
  - Prices
  - Dates/Times
  - Hashtags

**Post 1 (X / Twitter)**

> "Groceries are wild lately. $70 and somehow still nothing to cook 😩"

**Post 2 (Instagram story caption)**

> "Made a salad for lunch today… then ordered fries at 4pm 🤷‍♀️"

## Sentiment analysis

- Classify posts into Negative/Positive/Neutral

"Fast food is cheap but leaves me feeling guilty. Taste is okay."

"The grocery delivery app is easy to use, but the delivery was late."

But we should really talk about Aspect-Based Sentiment Analysis

| Text |
| --- |
| "Groceries are ridiculous lately. $80 and I still need dinner ideas. #groceries #inflation" |
| "Salad for lunch today 🥗 small wins. #healthyeating #lunch" |
| "I want to cook more but I'm exhausted. #homecooking #busy" |
| "Meal prep looks great until day 3... #mealprep #reallife" |
| "Fast food is cheaper than cooking right now... #fastfood #costofliving" |
| "Trying the new vegan restaurant. Love the taste! #vegan #yum" |
| "My coffee machine broke. Now mornings are stressful. #coffee #fail" |
| "I finally organized my pantry, so much easier to cook now! #organization #mealprep" |
| "Delivery was late again... but food was still hot. #takeout #frustration" |
| "Cooking with kids is fun but messy! #family #cooking" |
| "I'm saving money by buying in bulk. #groceries #budget" |
| "Trying intermittent fasting, mornings are rough though. #fasting #health" |

## 1. Automatically identify the Aspects

Example aspects for eating habits posts:

- Food cost / budget
- Food quality / taste
- Convenience / time
- Cooking / meal prep
- Service / delivery
- Kitchen / tools
- Health / diet
- Experience / stress / enjoyment

| PostID | Food cost | Food quality | Convenience | Cooking/Meal prep | Service/Delivery | Kitchen/Tools | Health/Diet | Experience/Stress |
|---|---|---|---|---|---|---|---|---|
| SM01 | Negative | Neutral | Negative | Negative | NA | NA | NA | Negative |
| SM02 | NA | Positive | Positive | NA | NA | NA | Positive | Positive |
| SM03 | NA | NA | Negative | Positive | NA | NA | NA | Negative |
| SM04 | NA | Positive | Neutral | Positive | NA | NA | NA | Neutral |
| SM05 | Positive | Neutral | Positive | NA | NA | NA | Negative | Negative |
| SM06 | NA | Positive | NA | NA | NA | NA | Positive | Positive |
| SM07 | NA | NA | NA | NA | NA | Negative | NA | Negative |
| SM08 | NA | NA | Positive | Positive | NA | Positive | NA | Positive |
| SM09 | NA | Positive | NA | NA | Negative | NA | NA | Neutral |
| SM10 | NA | Positive | NA | Positive | NA | NA | NA | Positive |
| SM11 | Positive | NA | Positive | NA | NA | NA | NA | |
| SM12 | NA | NA | Negative | NA | NA | NA | Positive | |

Sentiment classification per aspect + structuring

# Transactions

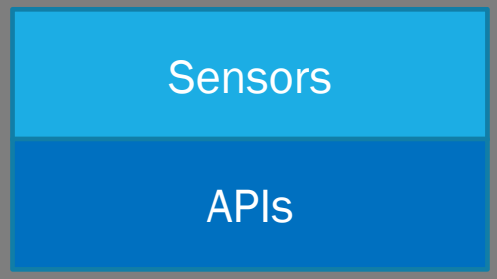
Case Study
Eating Habits

| TransactionID | Date | StoreType | ItemDescription | Category | Qty | TotalCost |
|---|---|---|---|---|---|---|
| 845210 | 2025-02-28 | Grocery | Rotisserie Chicken | Prepared Foods | 1 | 9.99 |
| 845211 | 2025-02-28 | Grocery | White Bread (Store) | Bakery | 1 | 2.49 |
| 845212 | 2025-02-28 | Grocery | Bananas | Produce | 6 | 1.86 |
| 845490 | 2025-03-01 | Fast Food | Large Fries | Side | 1 | 4.29 |
| 846102 | 2025-03-02 | Delivery App | Chicken Shawarma Plate | Restaurant Meal | 1 | 16.75 |
| 846330 | 2025-03-03 | Grocery | Frozen Lasagna | Frozen Meals | 1 | 6.99 |
| 846998 | 2025-03-04 | Coffee Shop | Latte (Medium) | Beverage | 1 | 5.25 |

# Comparison / Integration

| | Social media | Transactions |
|---|---|---|
| Type of data | Text (unstructured) and more | Categorical/Numerical (structured) |
| Captures | Opinions (often extreme ones) | Actual purchases |
| Best for | Understanding motivation (do a selection and study) | Measure behavior |
| Required tools | Text analysis (imperfect) | Validation |

# Case Study
## *Energy Consumption*



Case Study
Energy Consumption

# Sensors

## Sensor Types Included

1. Smart electricity meter
2. Smart thermostat
3. Indoor temperature sensor
4. Occupancy / motion sensor

**Electricity Smart Meter (Whole-Home)**

| timestamp | meter_id | power_kW |
|---|---|---|
| 2025-02-12 06:00:00 | MTR_88421 | 0.8 |
| 2025-02-12 06:15:00 | MTR_88421 | 1.1 |
| 2025-02-12 06:30:00 | MTR_88421 | 1.6 |
| 2025-02-12 06:45:00 | MTR_88421 | 2.3 |
| 2025-02-12 07:00:00 | MTR_88421 | 3.0 |

**Smart Thermostat**

| timestamp | thermostat_id | setpoint_C | indoor_temp_C | hvac_mode |
|---|---|---|---|---|
| 2025-02-12 05:50:12 | THM_33210 | 21.0 | 19.4 | heating |
| 2025-02-12 06:10:08 | THM_33210 | 21.0 | 19.9 | heating |
| 2025-02-12 06:28:44 | THM_33210 | 22.0 | 20.3 | heating |
| 2025-02-12 06:47:01 | THM_33210 | 22.0 | 20.9 | heating |
| 2025-02-12 07:15:33 | THM_33210 | 20.0 | 21.2 | off |

## Sensor Types Included

1. Smart electricity meter
2. Smart thermostat
3. Indoor temperature sensor
4. Occupancy / motion sensor

### Indoor Temperature Sensors (Room-Level)

| timestamp | temp_sensor_id | room | temperature_C |
|---|---|---|---|
| 2025-02-12 06:00:00 | TMP_LR_091 | LivingRoom | 19.6 |
| 2025-02-12 06:05:00 | TMP_LR_091 | LivingRoom | 19.8 |
| 2025-02-12 06:10:00 | TMP_LR_091 | LivingRoom | 20.1 |
| 2025-02-12 06:15:00 | TMP_LR_091 | LivingRoom | 20.4 |
| 2025-02-12 06:20:00 | TMP_LR_091 | LivingRoom | 20.7 |

### Occupancy / Motion Sensors

| timestamp | motion_sensor_id | room | event_type | occupancy_detected |
|---|---|---|---|---|
| 2025-02-12 06:02:11 | OCC_BED_7712 | Bedroom | motion | TRUE |
| 2025-02-12 06:18:47 | OCC_KIT_8894 | Kitchen | motion | TRUE |
| 2025-02-12 06:45:03 | OCC_BTH_3341 | Bathroom | motion | TRUE |
| 2025-02-12 07:32:56 | OCC_LR_4502 | LivingRoom | motion | TRUE |
| 2025-02-12 08:15:29 | OCC_LR_4502 | LivingRoom | no_motion | FALSE |

# APIs

Case Study
Energy Consumption

## Weather APIs

| API Example | Data Provided | Example |
|---|---|---|
| OpenWeatherMap / WeatherAPI | Temperature, humidity, wind speed, precipitation, conditions | Temp_C: -15.1, Humidity: 40%, Wind_kmh: 22, Precip_mm: 0.2, Condition: Snow |
| Weather.gov (US NOAA) | Hourly temperature, precipitation, storm alerts | Temp_C: 2, Precip_mm: 5, Snow_mm: 0 |
| Meteostat | Historical weather for trend analysis | Average daily temp, max/min, precipitation |

Various APIs

1. Weather APIs
2. Electricity Price APIs
3. Spatial (positioning) APIs
4. Calendar (events) APIs

Various APIs

1. Weather APIs
2. Electricity Price APIs
3. Spatial (positioning) APIs
4. Calendar (events) APIs

## Electricity Price APIs

| API Example | Data Provided | Example |
|---|---|---|
| IESO (Ontario) / EIA (US) | Hourly market price of electricity, demand forecasts | Price: 0.12 CAD/kWh, Demand: 12,500 MW |
| Nord Pool | Day-ahead market prices (Europe) | Price: €0.095/kWh, Forecast demand: 4,300 MW |

## Spatial (positioning) APIs

| API Example | Data Provided | Example |
|---|---|---|
| Google Maps / OpenStreetMap | Building footprint, land use, nearby facilities | Residential, commercial, industrial density |
| City of Ottawa Open Data | Street, building type, energy efficiency rating | Building type: Apartment, Year built: 1980, Energy rating: B |

## Calendar (events) APIs

| API Example | Data Provided | Example |
|---|---|---|
| Google Calendar API / Public Holiday APIs | National holidays, school breaks, daylight savings | Date: 2025-02-17, Holiday: Family Day |
| Event APIs (Ticketmaster, local city) | Major public events, sports, concerts | Event: Hockey Game, Date: 2025-02-10 |

Various APIs

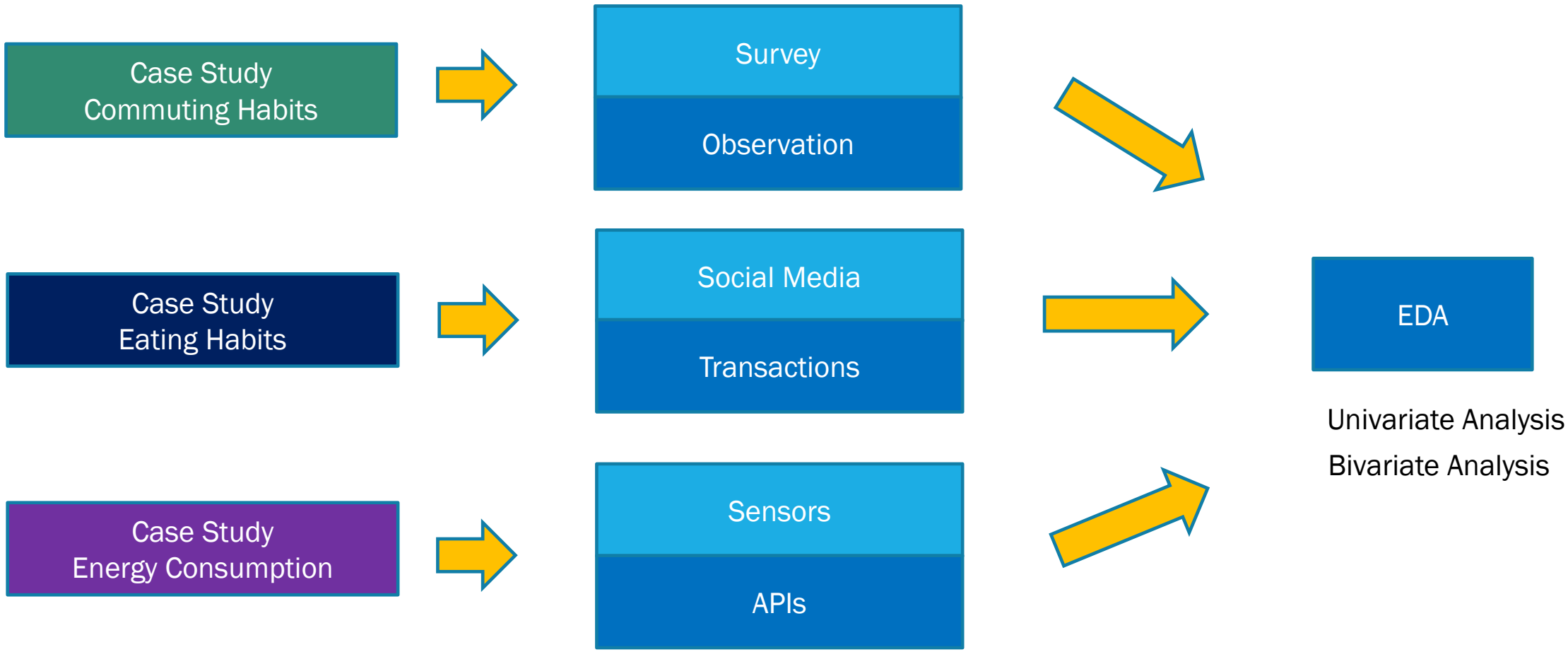
1. Weather APIs
2. Electricity Price APIs
3. Spatial (positioning) APIs
4. Calendar (events) APIs

## Comparison / Integration

| | Sensors | APIs |
|---|---|---|
| What is measured | Energy usage, user actions, user behaviors | External conditions (weather, location, events) |
| Source | Physical devices in-home | External data providers (APIs) |
| Temporal coverage | Continuous (unless sensor defect) | APIs availability |
| Spatial coverage | Individual household | City, region |

Case studies:

Focus on two collection methods.



# DATA COLLECTION

- Decision Factors for Data Collection Method Selection
- Three comparative case studies