

1. Craft a data mining problem statement based on a pipeline that you have created in the previous session;

→ Practical motivation and sample collection

- ◆ Let's highlight first the issue. The problem is the observability of the system while the user has an active session in the application. Visibility into the customer journey allows the company to quickly resolve issues and add more value to the digital experience. Because poor experience impacting the company's bottom line in ways might not even realize.

- **Is the problem related to data?**

- Yes, this type of problem is related to it. Data mining will help to resolve the issue by using intelligent methods in discovering useful patterns.

- **What type of data can be used to solve the problem?**

- I have listed the classification measurements that will serve as an input to the process of knowledge discovery.

- ◆ User detailed information (name, location, version, platform, etc)
- ◆ User application snapshot (current user view)
- ◆ Client application logs either warning, useful, errors (including API responses and request payloads)
- ◆ User rage clicks (user frustration)
- ◆ Segment health (events per session, session length, active time)
- ◆ Loading time (where is the slowest page)

- **Is the data highly available?**

- Yes, it will be as long as the customers are active within the application. Segment health is an important part of the tool because It distinguishes the session playlist, user trends, health, top users, and many more.

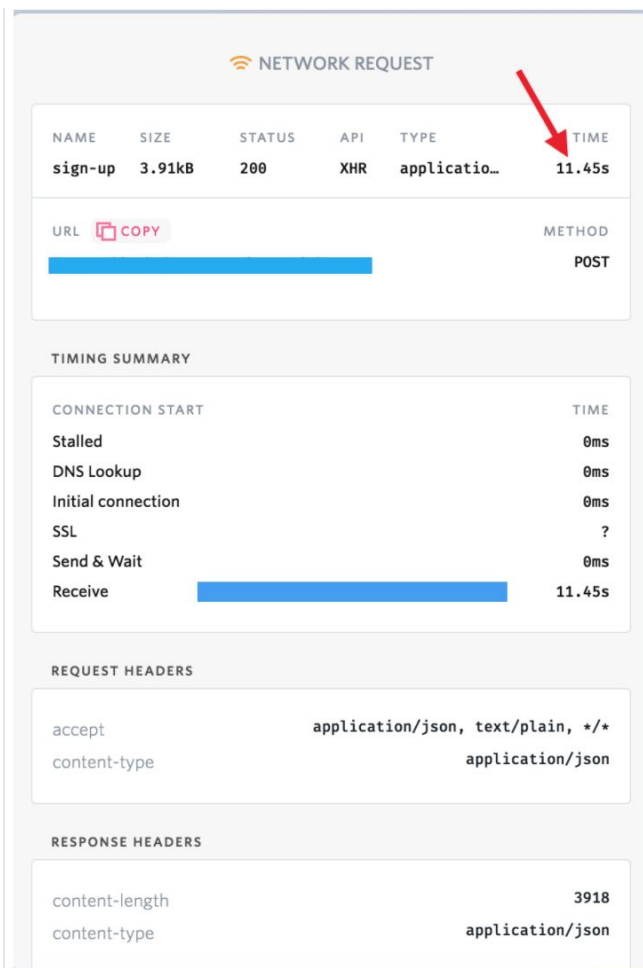
→ Problem formulation and data preparation

◆ Should network requests be considered?

- Yes, network integration must have visibility too. It is possible to capture the snapshot from the browser console.

The image below is a mock-up describing network details.

- Name
- Size
- Status
- API
- Type
- Timing summary



◆ **What is the format of data?**

- The data format must follow the standard schema model. It should be
 - Flatten key-value pair format
 - Lists of the item whenever possible
 - Payload accuracy
 - Cross-out duplicate values

→Statistical description and exploratory analysis

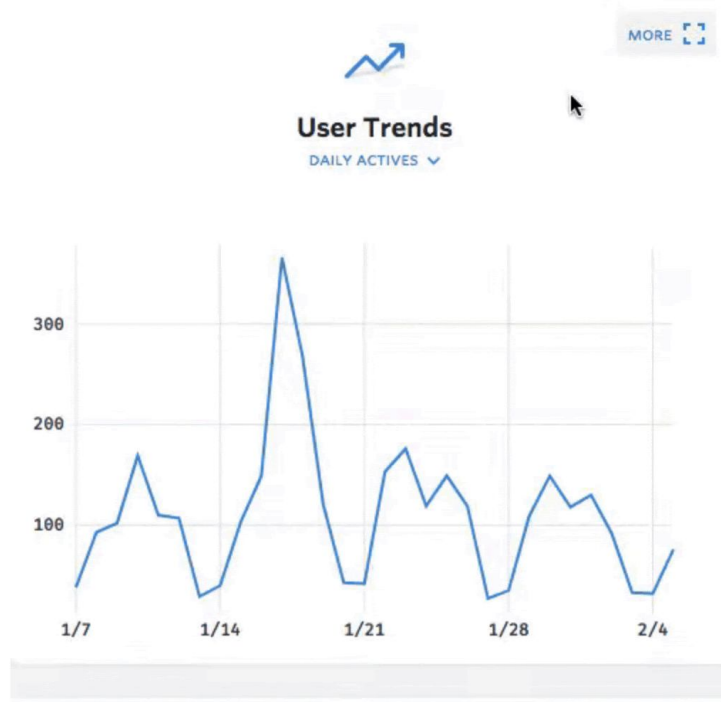
◆ **How can we clearly express the data with some representations? What statistical insights are relevant?**

- Since we're dealing with a massive amount of information, it is suggested to categorize and group them before doing exploratory analysis.
 - System logs
 - ◆ API response and request
 - ◆ Back end system health
 - ◆ Database performance
 - User information
 - ◆ Customer session
 - ◆ Current view snapshot
 - ◆ Trends
 - ◆ Top users
 - Client-side application performance
 - ◆ Client application logs
 - ◆ Speed (page loading time)
 - ◆ UX information (rage clicks, dead clicks, etc)
 - ◆ Network requests

→Pattern recognition and analytical visualization

◆ Exploratory analysis and visualization

- The first step in this procedure is to understand the data by having an initial investigation. The goal is to gather as many insights as possible. The image below displays user sessions in a line graph. What is this all about? User trends help people to analyze active users in the application. It can be daily, weekly, monthly sessions. It is a good practice to understand the data first before anything else.









→Machine learning and algorithmic optimization

◆ Machine learning and algorithmic optimization

- Now all datasets are complete and ready to utilize, the next step is to create predictions using a machine learning technology. An example is, how many opportunities loss from the user's next purchase?
- The image below is a table pertaining to the number of issues (client-side performance, system issue) experienced by the customer and impacting in the loss of conversions (revenue).

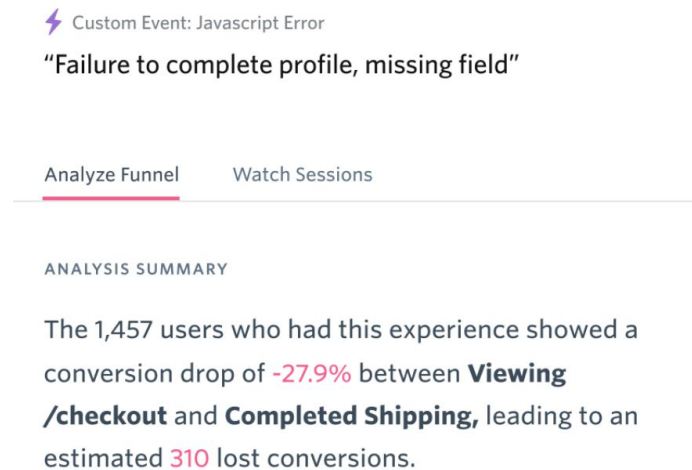
Opportunities to improve Checkout

Customer Experience		Affected User Sessions	Conversion Impact	Lost Conversions ▾
	Custom Event Javascript Error · message · "Failure to complete profile, missing field"	4,626	-74.27%	2,895
	Page Load Performance · 6+ seconds /cart	13,174	-23.19%	1,454
	Error Click Please enter a unique mobile phone number.	1,984	-86.42%	1,069
	Dead Click Update your shipping address	2,009	-83.72%	1,048
	Rage click Place Order	19,515	-7.22%	867
	Watched Element .errorDialog · "2034AKX0"	120,068	-2.48%	635

→Statistical inference and information presentation

◆ Analysis of summary and conclusion

- It can be difficult on which one should prioritize the improvement and fixes if customers struggling and experienced unnecessary problems. The image below tells about an actual error and analysis summary.



→Intelligence decision and ethical consideration

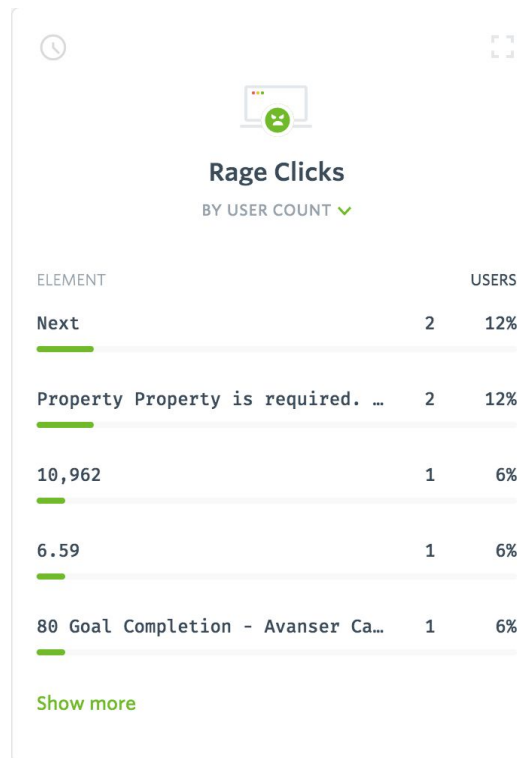
◆ User privacy

- By default and sensitive information like credit cards, government identifications, email is completely blocked from the monitoring tool.

2. Make distinctions of descriptive and inferential analysis

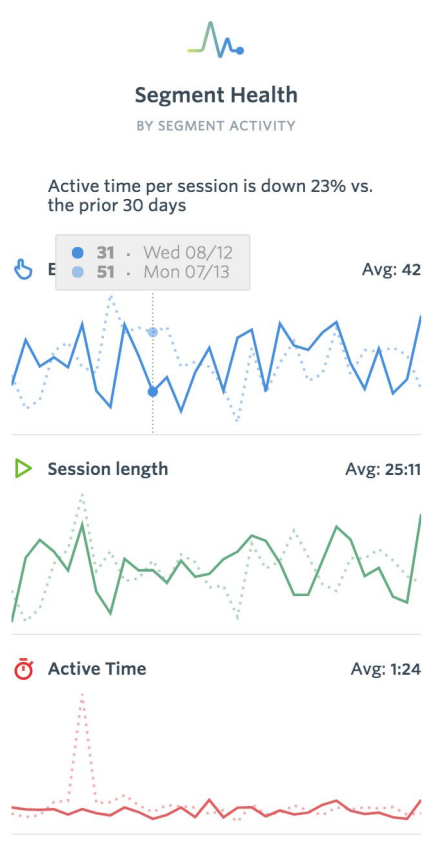
→ Descriptive analysis

- ◆ This is an example of a descriptive analysis because it answers the questions of what happened? And why? Data mining discovers a frustrating user experience events like rage clicks. This is happening when the user tried to click something that the system didn't respond quickly.



→Inferential analysis

- ◆ Unlike descriptive statistics, this data analysis can extend to a similar larger group and can be visually represented through graphic elements. A sample data is considered, studied, and analyzed. An example below is the actual segment health of the user's session.



3. Synthesize the steps in solving a data mining problem.
 - a. I have already explained this in detail at the number one question.