



Segmentifier: Interactive Refinement of Clickstream Data

Kimberly Dextras-Romagnino^{*†}, Tamara Munzner^{*}

^{*} University of British Columbia

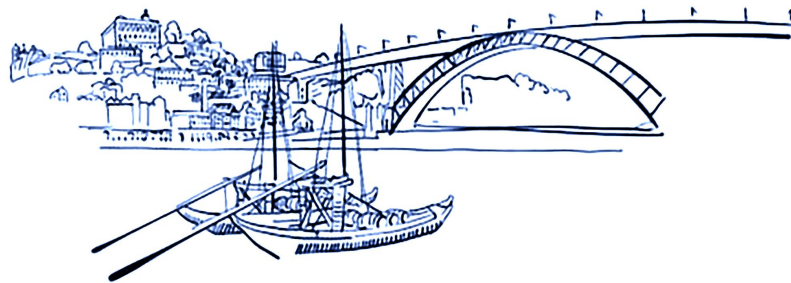
[†] Kabam



DESIGNING
for PEOPLE



KABAM



Introduction: *E-commerce*



E-commerce

- Build mobile apps for large e-commerce companies
- Understand the importance of good websites for revenue

E-commerce

- Build mobile apps for large e-commerce companies
- Understand the importance of good websites for revenue

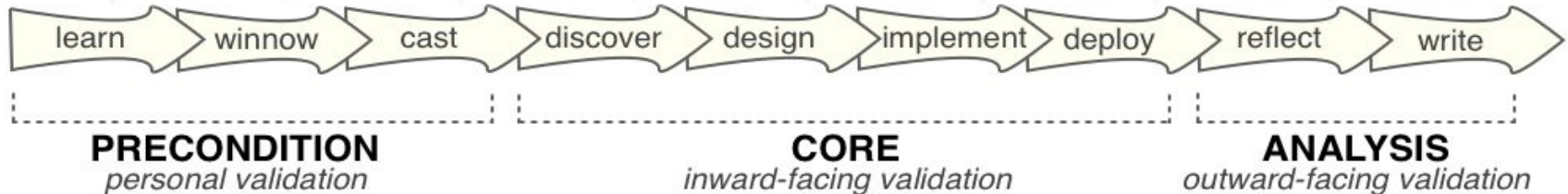
Goals

- **Increase traffic**
 - number of users on a site
- **Reduce abandonment**
 - number of users leaving the site
- **Increase consumer engagement**
 - time users spend on the site
 - chances that a user returns to the site
- **Increase conversion rate**
 - odds a user purchases

Process

Followed Design Study Methodology [Sedlmair 2012]:

- **Precondition Phase** (5 months) : interviews with 12 employees
- **Core Phase** (11 months): Iterative design and implementation
- **Analysis Phase** (3 months): Reflect and write



Research Contributions

- Thorough **characterization of task and data abstraction** for clickstream data analysis
 - **High-Level Segmentifier Analysis Model** abstracts iterative process
 - View, Refine, Record, Export, Abandon, Conclude
- **Segmentifier: novel analytics interface** for refining data segments and viewing characteristics before downstream fine-grained analysis
 - Rich set of views showing both *derived attributes* and *raw sequence details*
 - *Filtering* and *Partitioning* through visual queries
 - Quantitative attributes
 - Custom sequences of events aggregated according to a novel three-level hierarchy
 - Detailed glyph based *visual history* of the automatically recorded refinement process showing the provenance of each segment in terms of its analysis path
- Preliminary **evidence of utility** from:
 - *Usage Scenario* with real world data
 - *Case Study* with industry analyst

What are the **Data and Task Abstractions** for *Clickstream Data Analysis*?

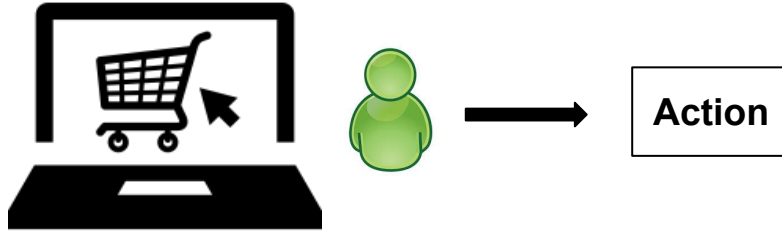
Clickstream Data

Clickstream Analysis Tasks

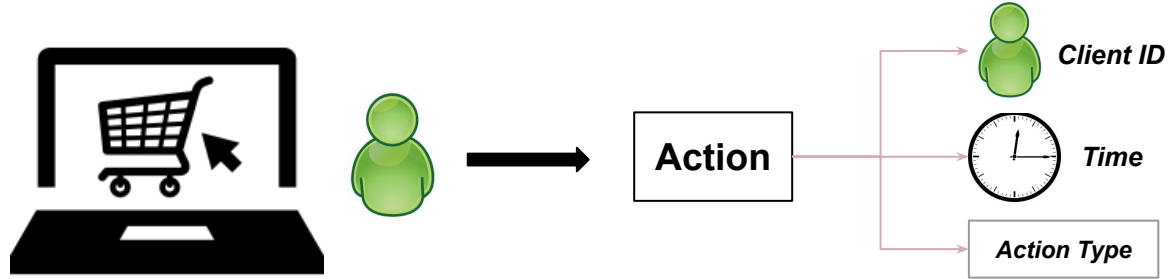
Segmentifier Analysis Model

What is ***Clickstream Data***?

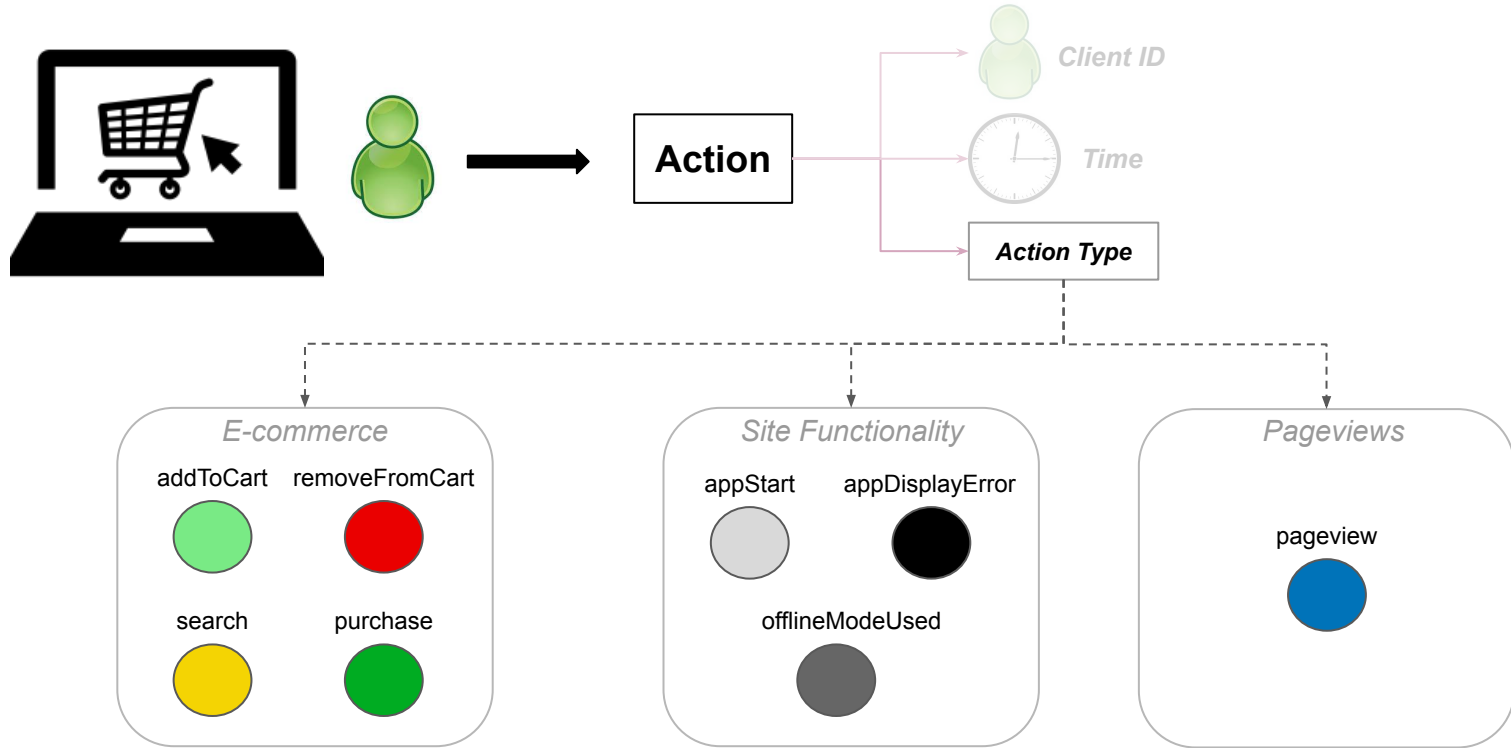
Data: *Actions*



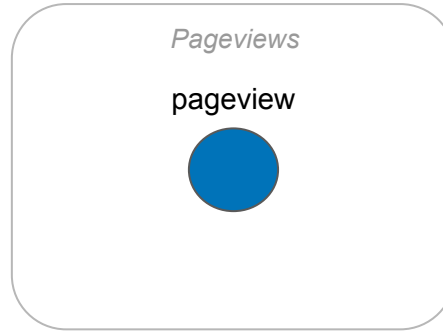
Data: *Action Attributes*



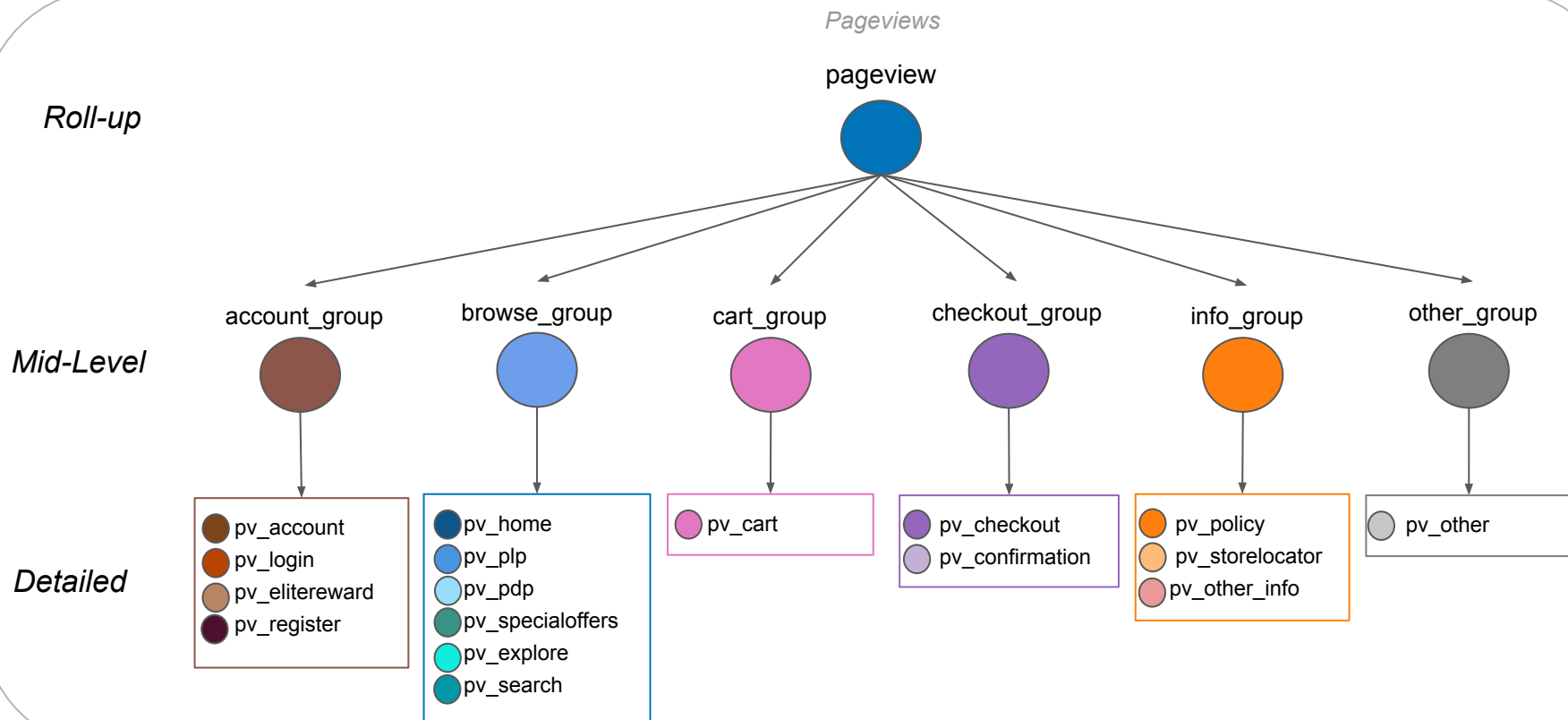
Data: *Action Types*



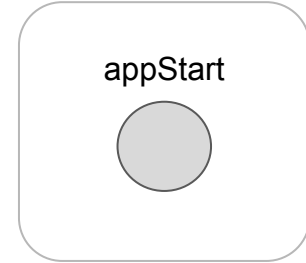
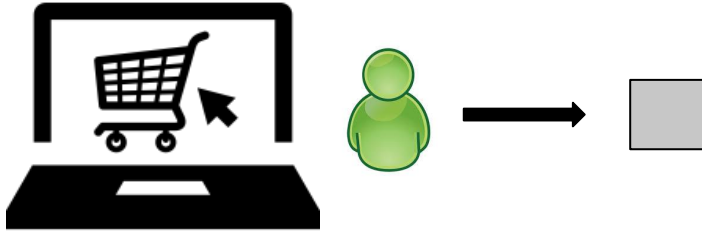
Action Hierarchy



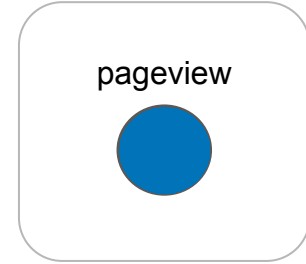
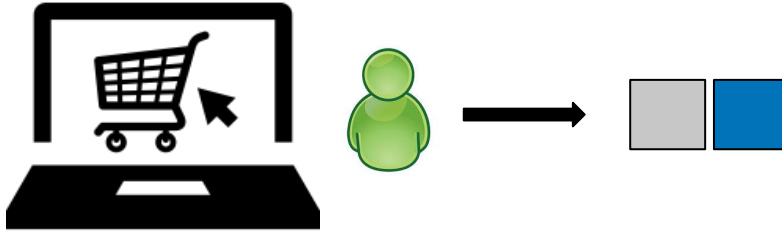
Action Hierarchy



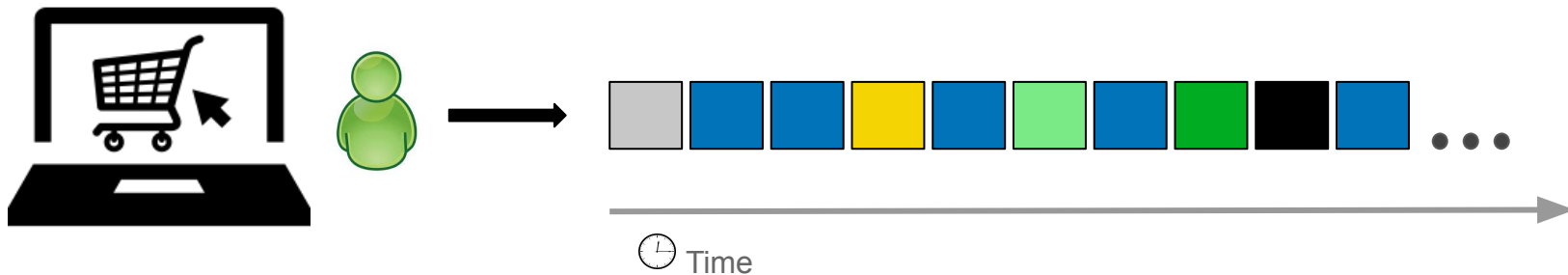
Data: *Sequences*



Data: *Sequences*

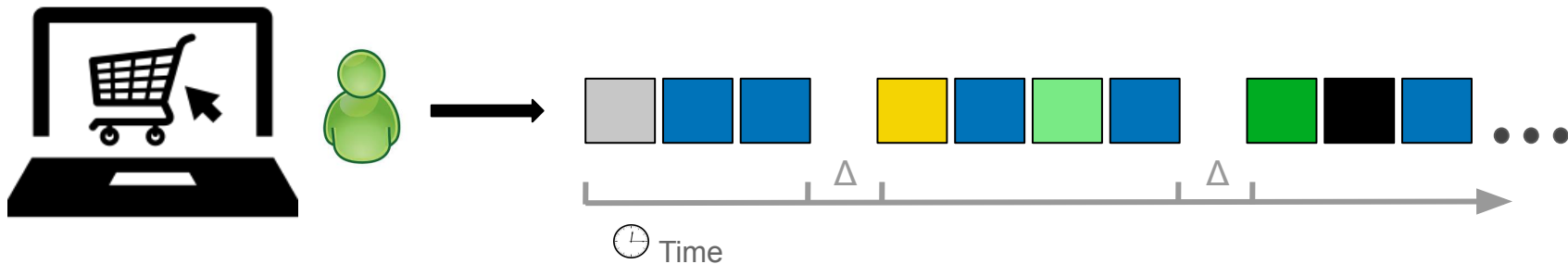


Data: *Client Sequences*



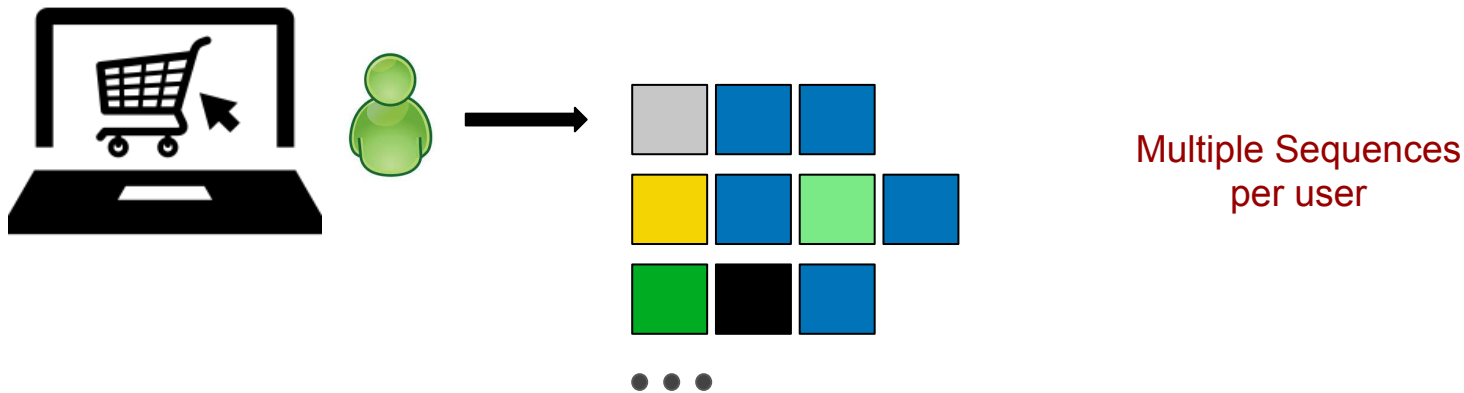
Client Sequences: all actions performed by a single user

Data: *Session Sequences*



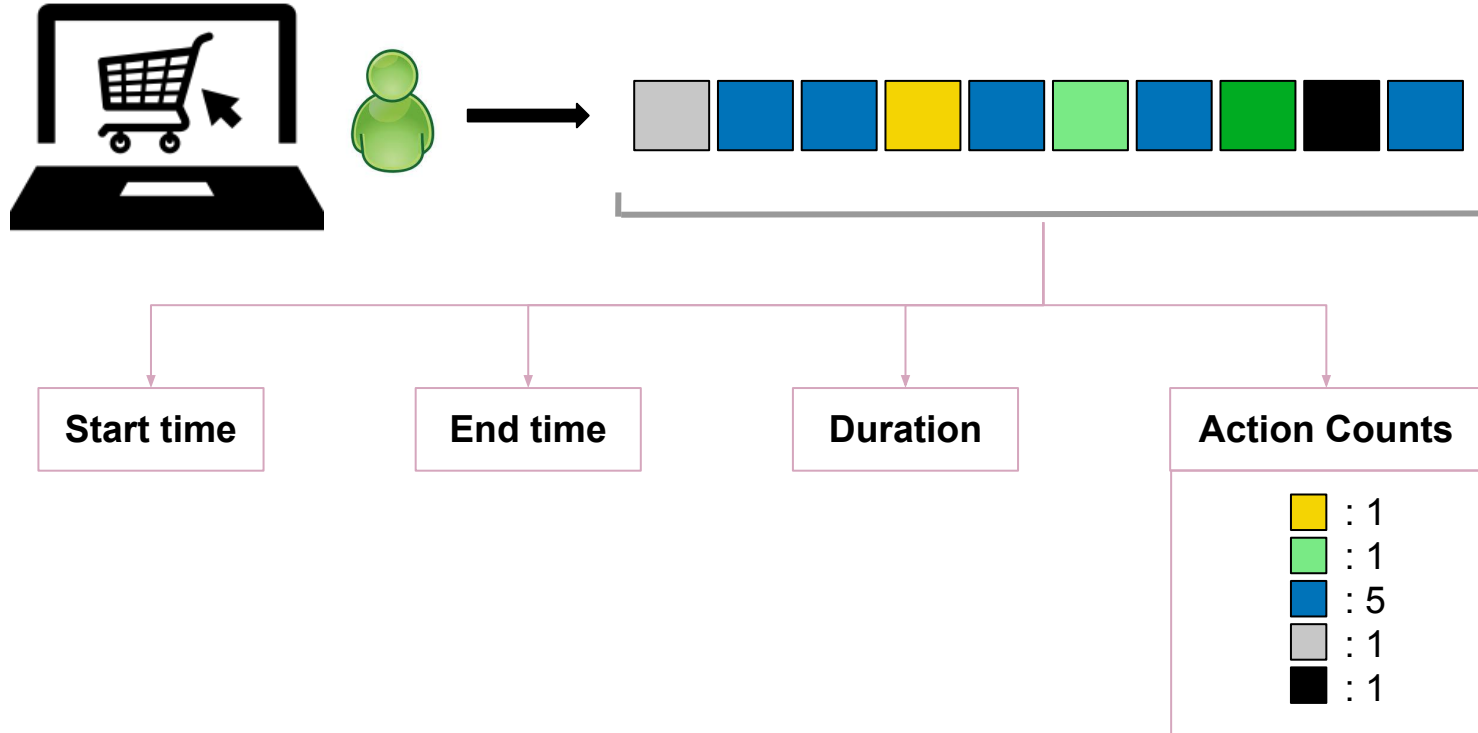
Session Sequences: all actions performed by a single user within a defined amount of time (Δ) from each other. Δ is usually 30 min.

Data: *Session Sequences*

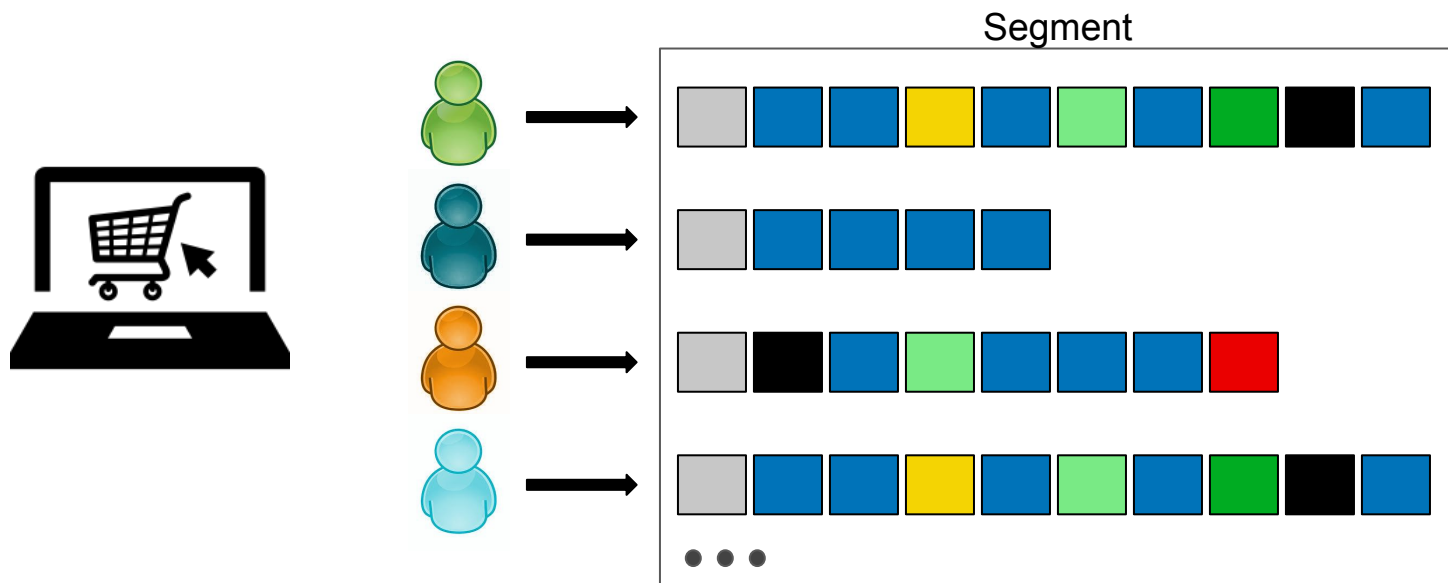


Session Sequences: all actions performed by a single user within a defined amount of time (Δ) from each other. Δ is usually 30 min.

Data: *Sequence Attributes*

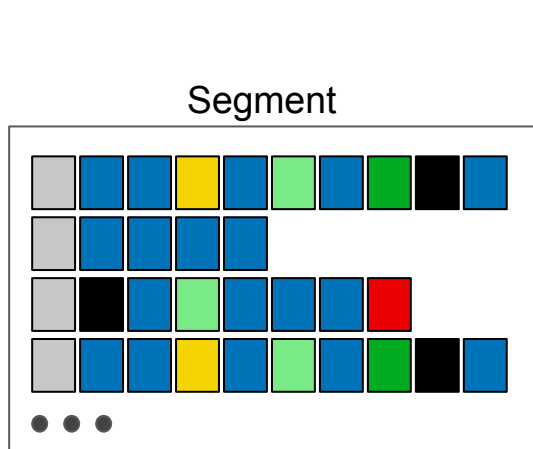


Data: *Segments*



Segment: any set of sequences

Data: *Segment Attributes*



Size

Counts of sequences:
Absolute, Relative

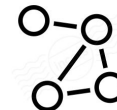
**Sequence
Related**

Sequence Distributions:
*Start Time, Duration, Action
Counts*

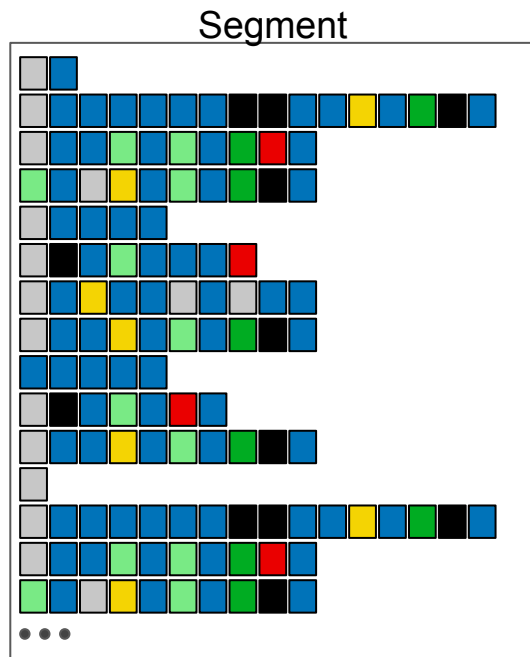


**Action
Related**

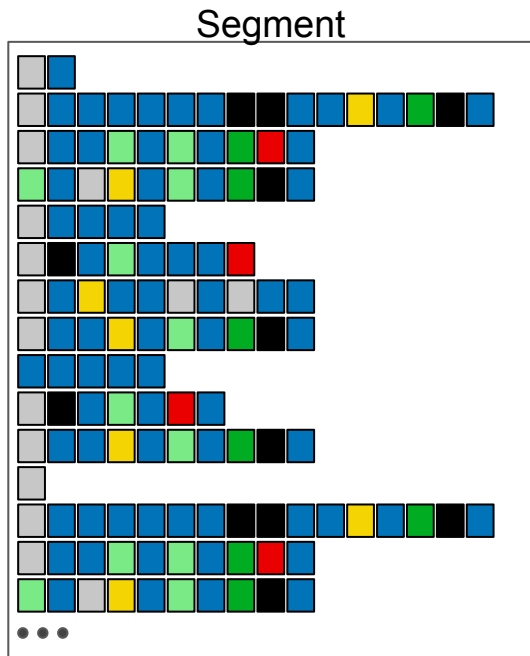
Action Distributions:
*Action Transitions:
action before, action after*



Real-world Clickstream Data

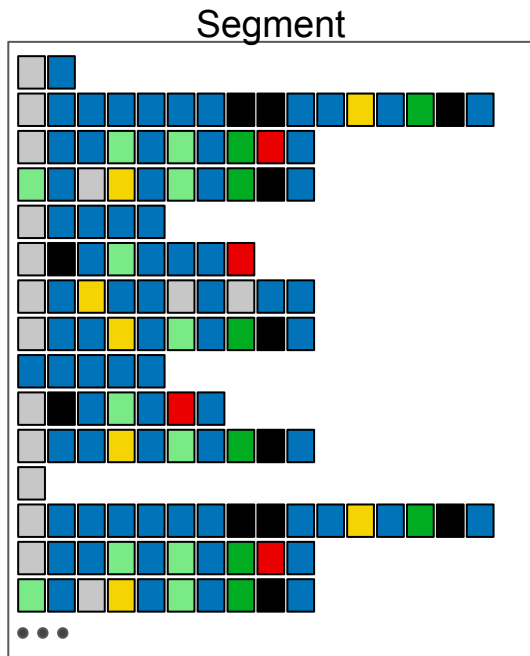


Real-world Clickstream Data



Scale is huge

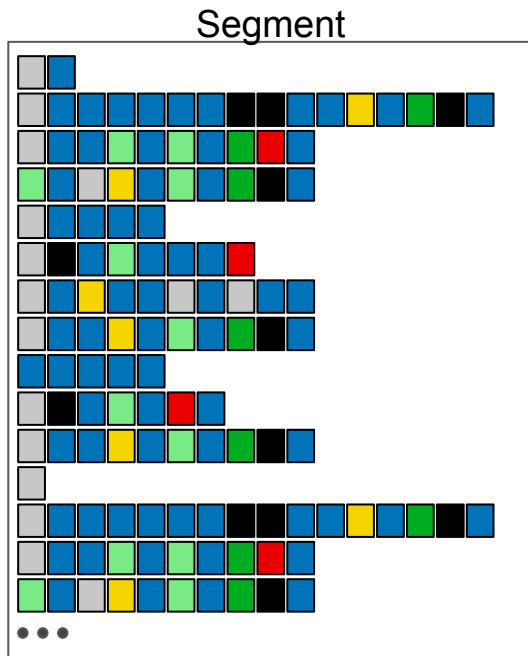
Real-world Clickstream Data



Scale is huge

Variability is high

Real-world Clickstream Data



Scale is huge

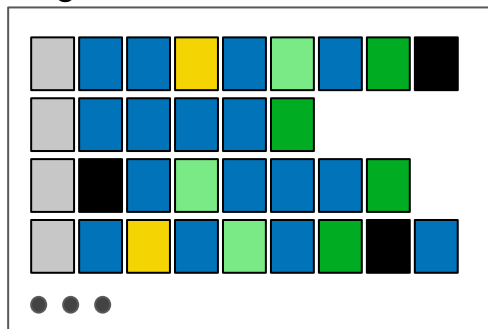
Variability is high

Most work **fails** when applied to real-world data.

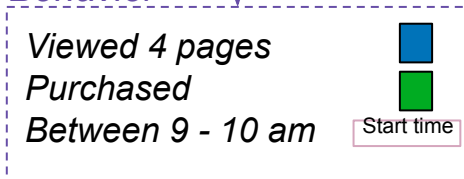
What are
Clickstream Data Analysis Tasks?

Tasks: Segment Behavior

Segment



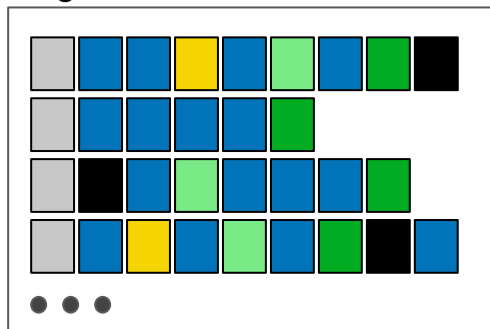
Behavior



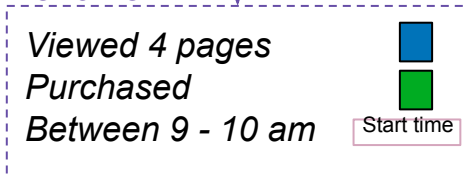
Behavior: set of attribute constraints

Tasks: Segment Behavior

Segment



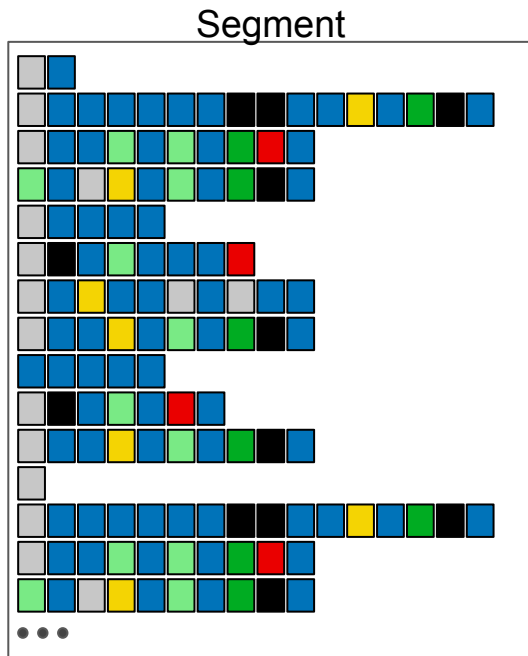
Behavior



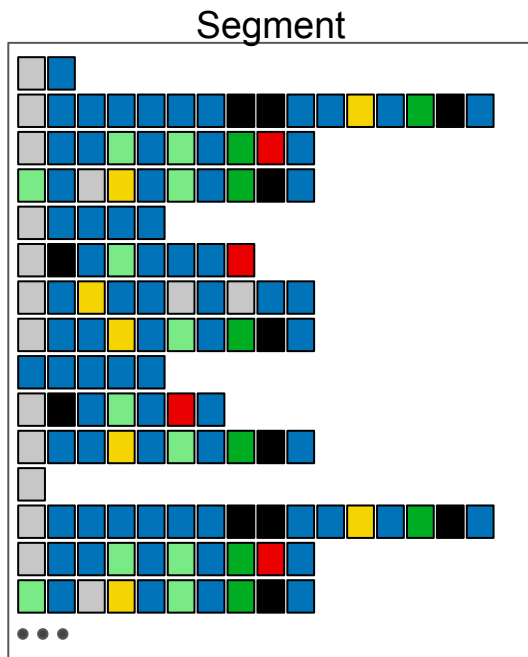
Behavior: set of attribute constraints

- **Expected**
Users add to cart before purchasing
- **Unexpected**
No purchases on a certain month
- **Favorable**
Purchased
- **Unfavorable**
Bounced

Tasks: Task Abstraction

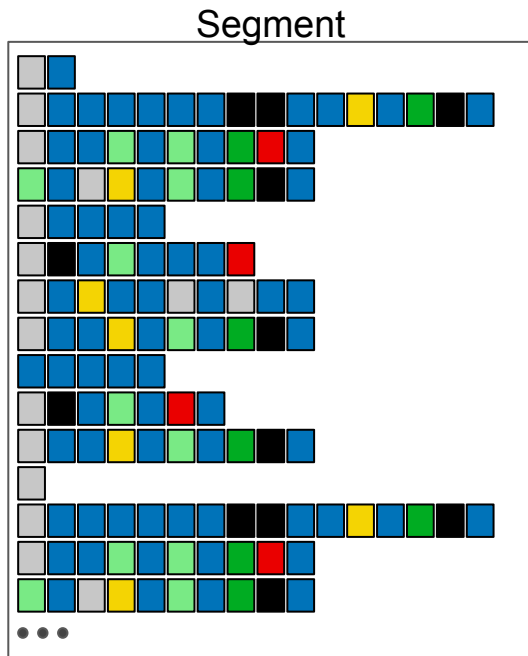


Tasks: Task Abstraction



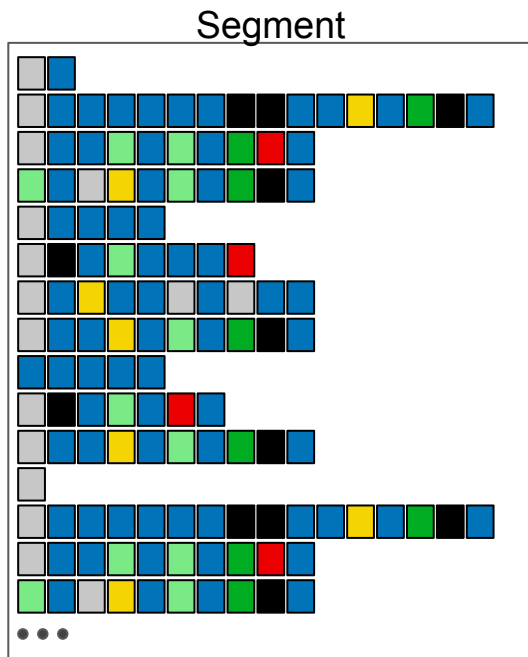
1. **Identify:** Find some set of sequences that constitutes interesting *behavior*

Tasks: Task Abstraction



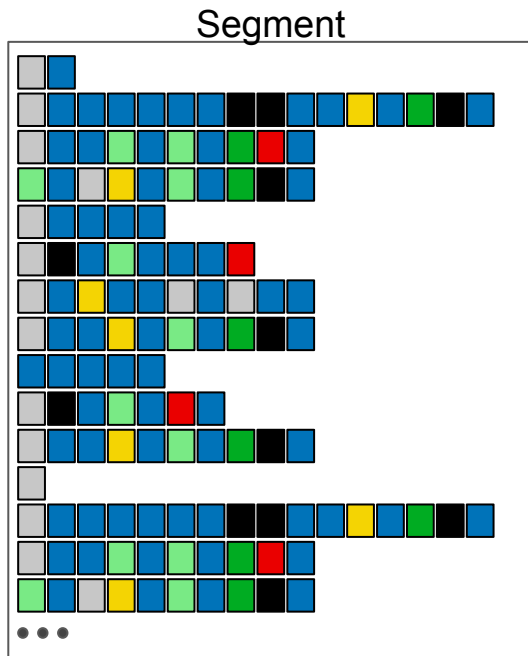
1. **Identify:** Find some set of sequences that constitutes interesting *behavior*
2. **Drilldown:** Distinguish more specific *behaviors* to further partition a segment previously defined by looser constraints

Tasks: Task Abstraction



1. **Identify:** Find some set of sequences that constitutes interesting *behavior*
2. **Drilldown:** Distinguish more specific *behaviors* to further partition a segment previously defined by looser constraints
3. **Frequency:** Determine how many sequences are in the segment defined by a *behavior*

Tasks: Task Abstraction



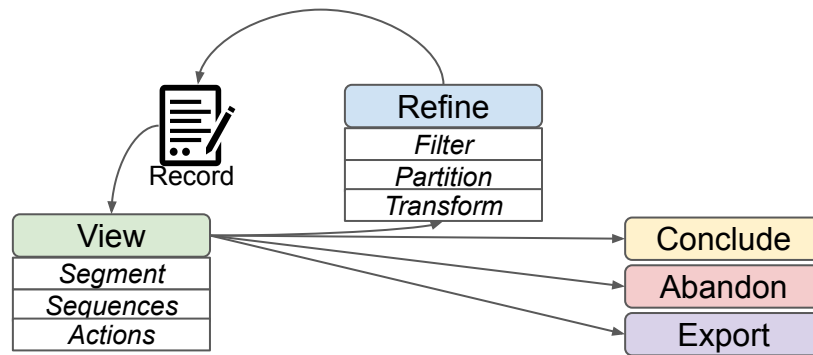
1. **Identify:** Find some set of sequences that constitutes interesting *behavior*
2. **Drilldown:** Distinguish more specific *behaviors* to further partition a segment previously defined by looser constraints
3. **Frequency:** Determine how many sequences are in the segment defined by *behavior*
4. **Ordering** within sequence: Match if one action subsequence occurs before (or after) another action subsequence in a sequence

High-Level Segmentifier Analysis Model

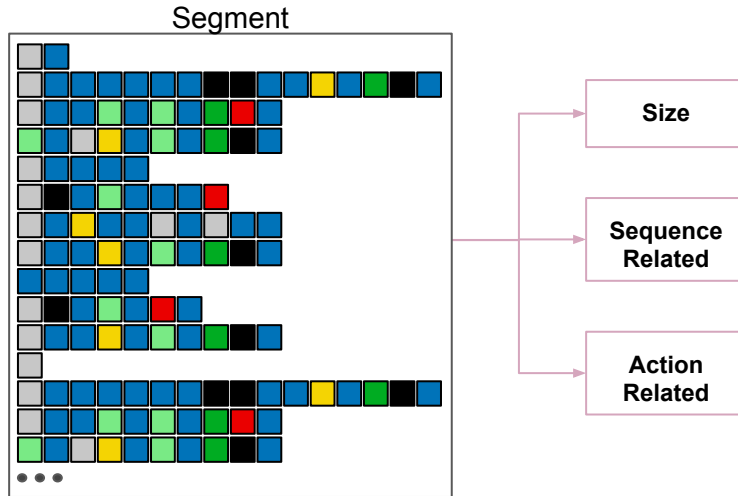
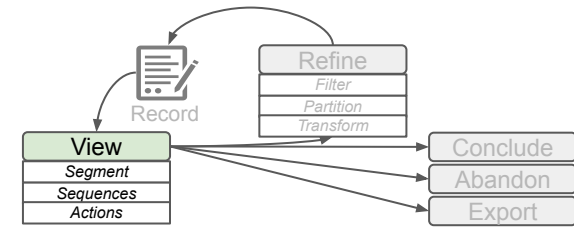
High-Level Segmentifier Analysis Model

General idea:

Combine domain knowledge with computational support to iteratively view and refine large, noisy clickstream segments into segments that lead to **actionable insights** or more effective **downstream analysis**

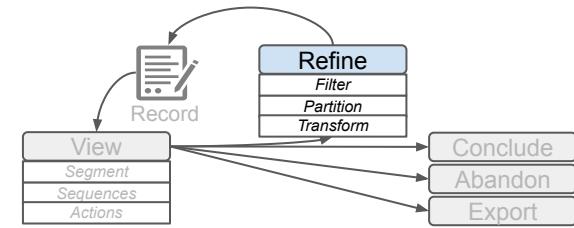
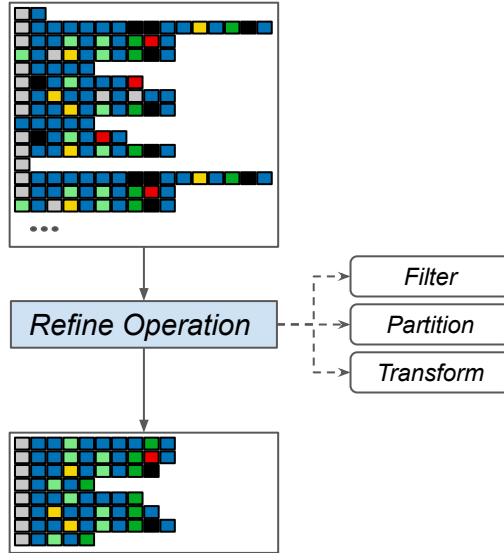


High-Level Segmentifier Analysis Model



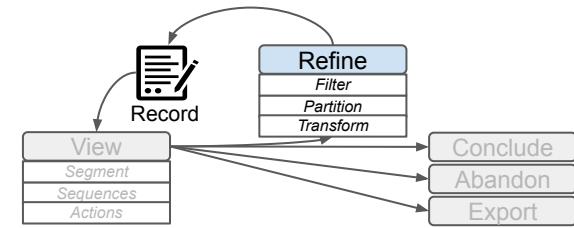
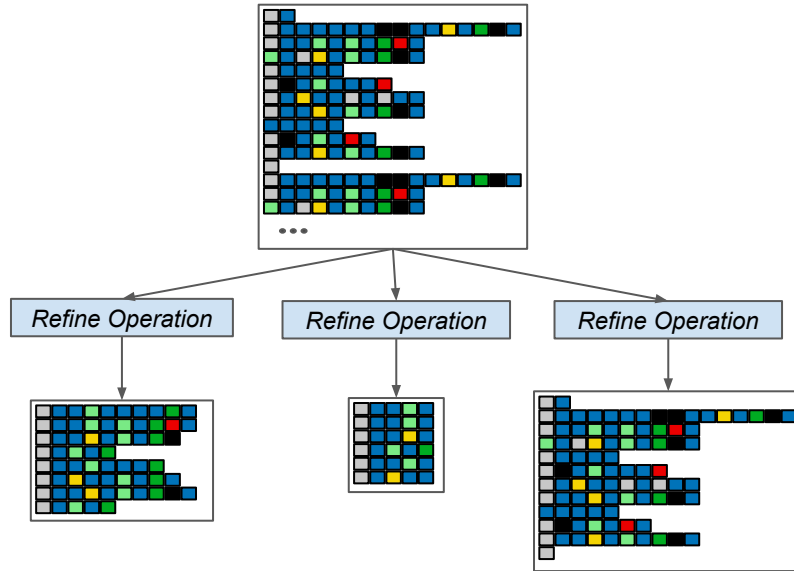
- Gives Insight into underlying data of segment
 - Action Attributes
 - Sequence Attributes
 - Segment Attributes
- Leads to:
 - Insights
 - New ways on how to *refine*
 - Whether segment should be *abandoned*
 - Whether segment should be *exported*

High-Level Segmentifier Analysis Model



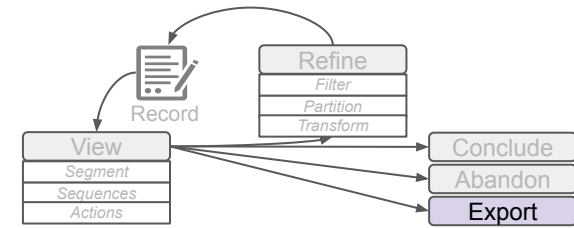
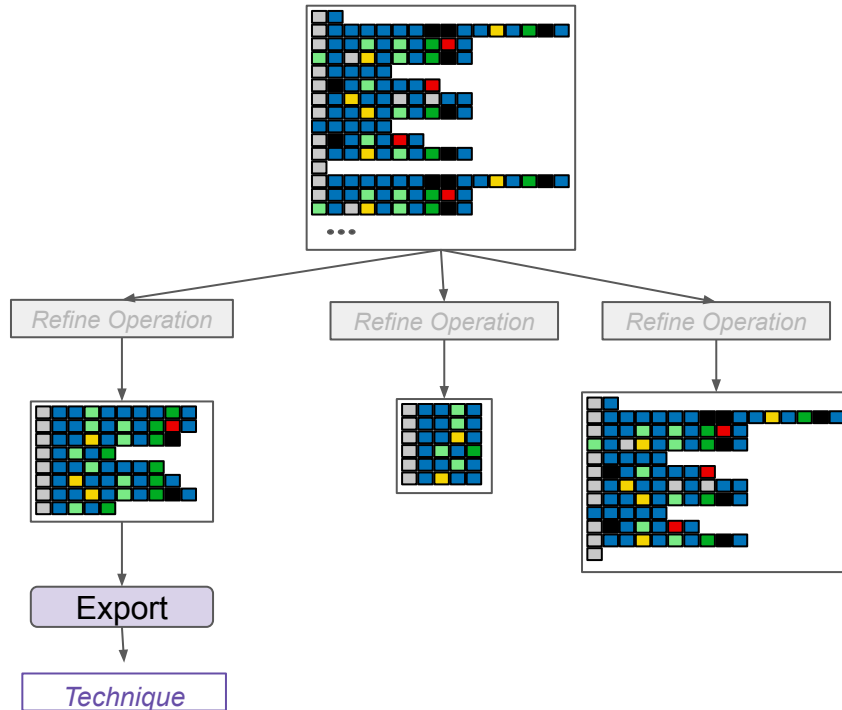
- Apply operation to create new segments
- Type of Refinements
 - *Filter*
 - *Partition*
 - *Transform*

High-Level Segmentifier Analysis Model



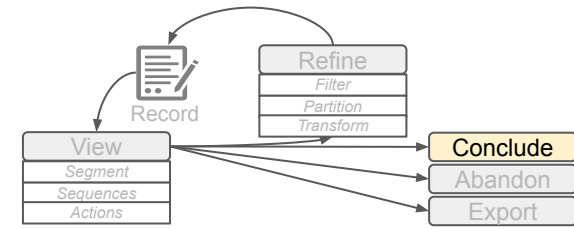
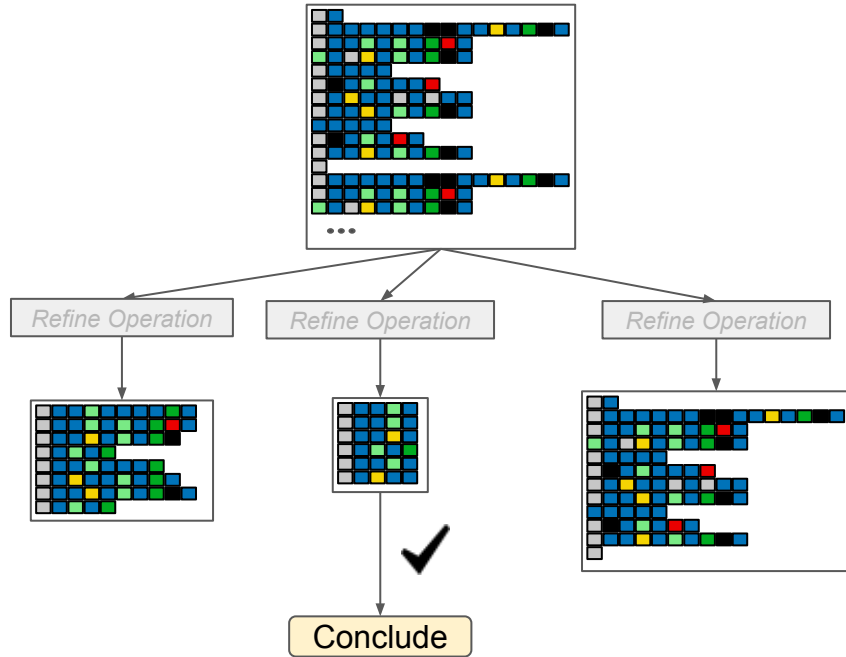
- Record all refinement steps automatically
- Keep track of questions asked and hypotheses tested
- Ability to create and view multiple segments from the same segment

High-Level Segmentifier Analysis Model



- Export refined segments for further downstream analysis, to more specific tools:
 - Pattern mining
 - Clustering

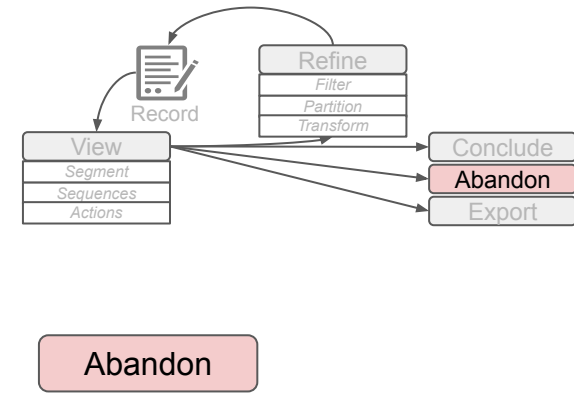
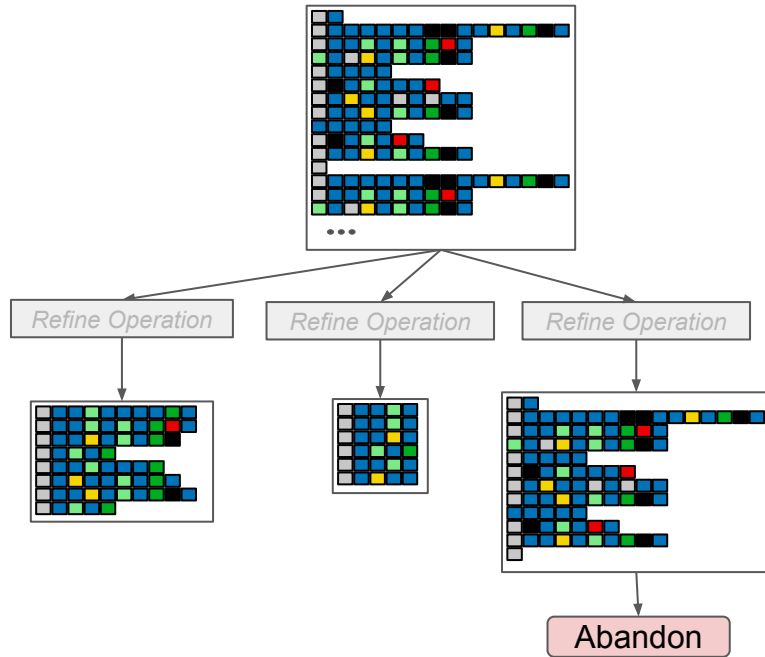
High-Level Segmentifier Analysis Model



Conclude

- Discover actionable insight by *viewing* segment

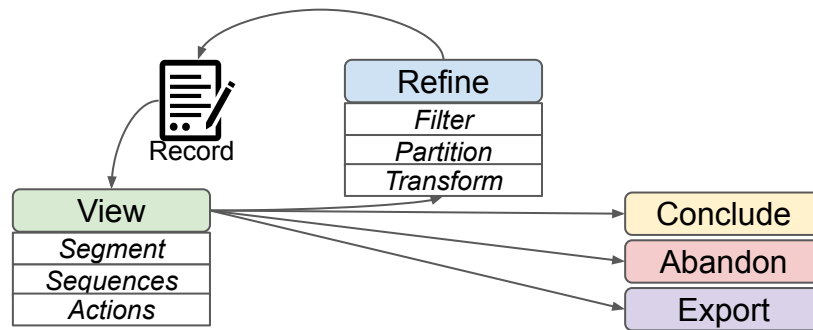
High-Level Segmentifier Analysis Model



- By *viewing* the segment, analyst *abandons* if:
 - No actionable insights
 - No further ways to *refine*
 - Not suitable for *export*

High-Level Segmentifier Analysis Model

- Take a *giant, noisy dataset* and refine it into *small, clean segments* appropriate for each *task*
- Bridge the gap between *real-world data* and other techniques
- Encapsulates the design rationale of **Segmentifier**



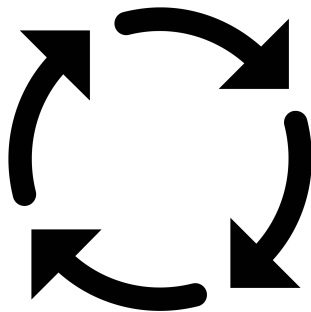
How to solve these goals with **Visual Analytics?**

Visual Analytics

Other Related Work

Our Framework

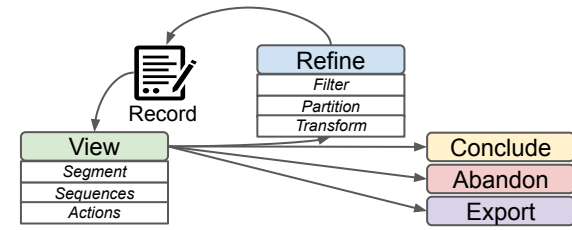
Why Visual Analytics?



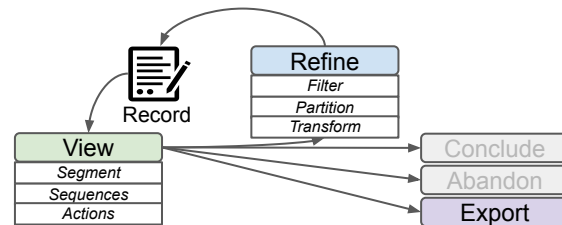
- Automation would be nice...
 - Put data in, actionable results appear
- ... but it is not realistic
 - Many possible questions, data-driven interplay between finding answers and generating new questions
- Human-in-the-loop visual data analysis
 - Integrate computing power of machine with intuition of domain experts

What **Visual Analytics Systems**
exist for **Clickstream Data Analysis**?

Related Work



Related Work



Export

Post-Export: Specific Techniques

- Clustering: [Wei 2012]. Pattern Mining: CoreFlow [Liu 2017], Frequence [Perer 2014]
- Require small, clean datasets

View

View Sequences: Event Sequence Visual Overviews

- CareFlow [Perer 2013]
- Limited ability to refine segments or view segment attributes

Refine

Refine: Visual Query Systems

- COQUITO [Krause 16], (s|qu)eries [Zgraggen 2015]
- No ability to view attributes

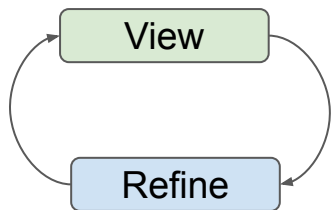


Record

Record: Graphical Histories

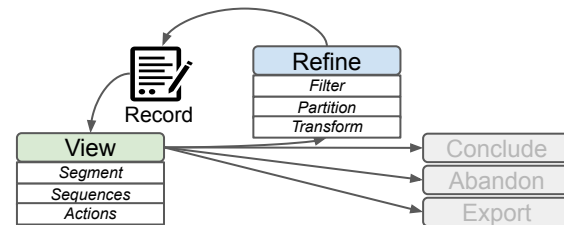
- Graphical histories help remember analysis path [Heer 2008]

Related Work



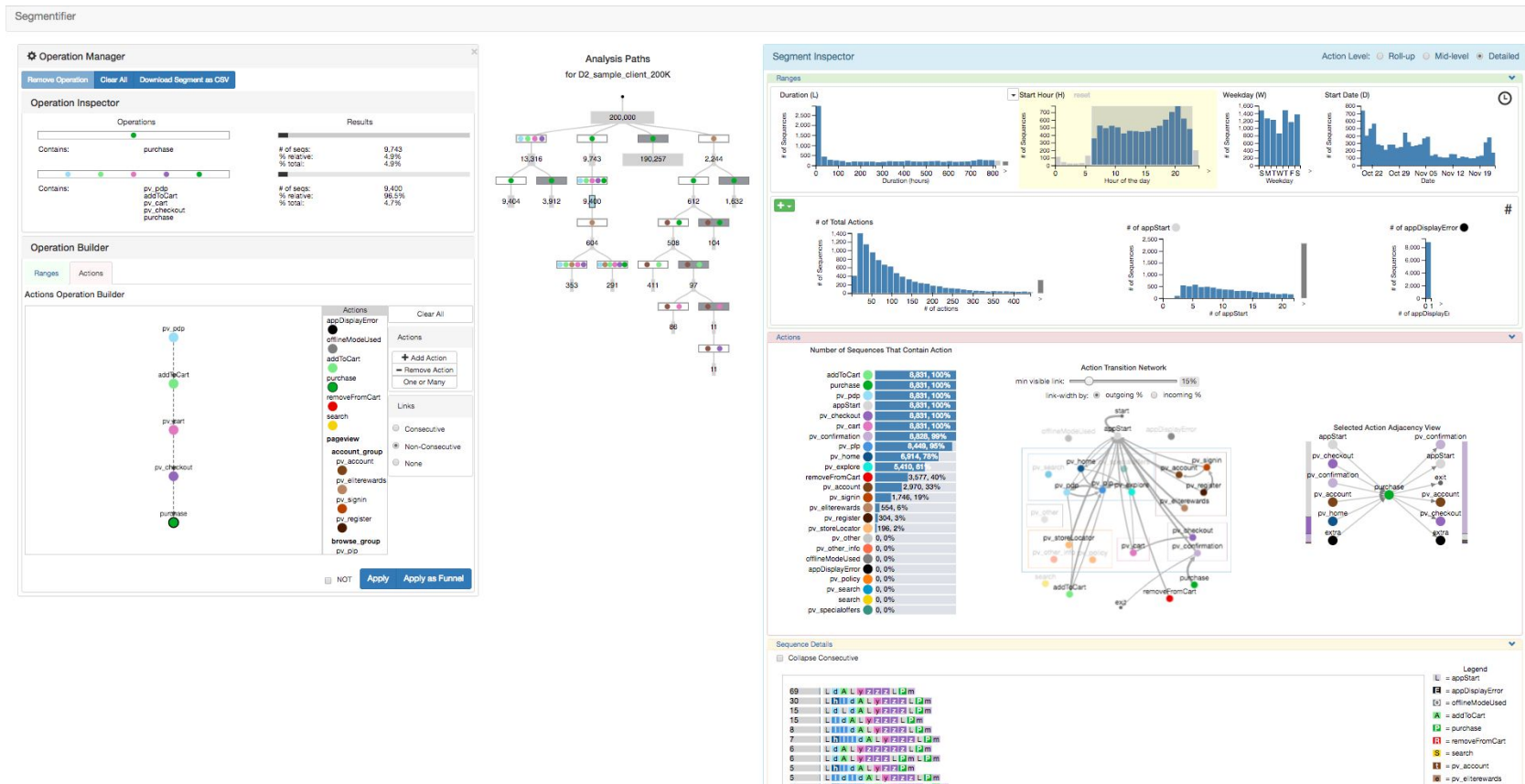
View and Refine: Filtering Sequences To Segments

- SessionViewer [Lam 2007], EventFlow [Munroe 2013] , EventPad [Cappers 2018]
- Lack of segment attributes
- Lack of ability to record analysis path
- Focus is on looking at the level of detail of the sequences which is unscalable

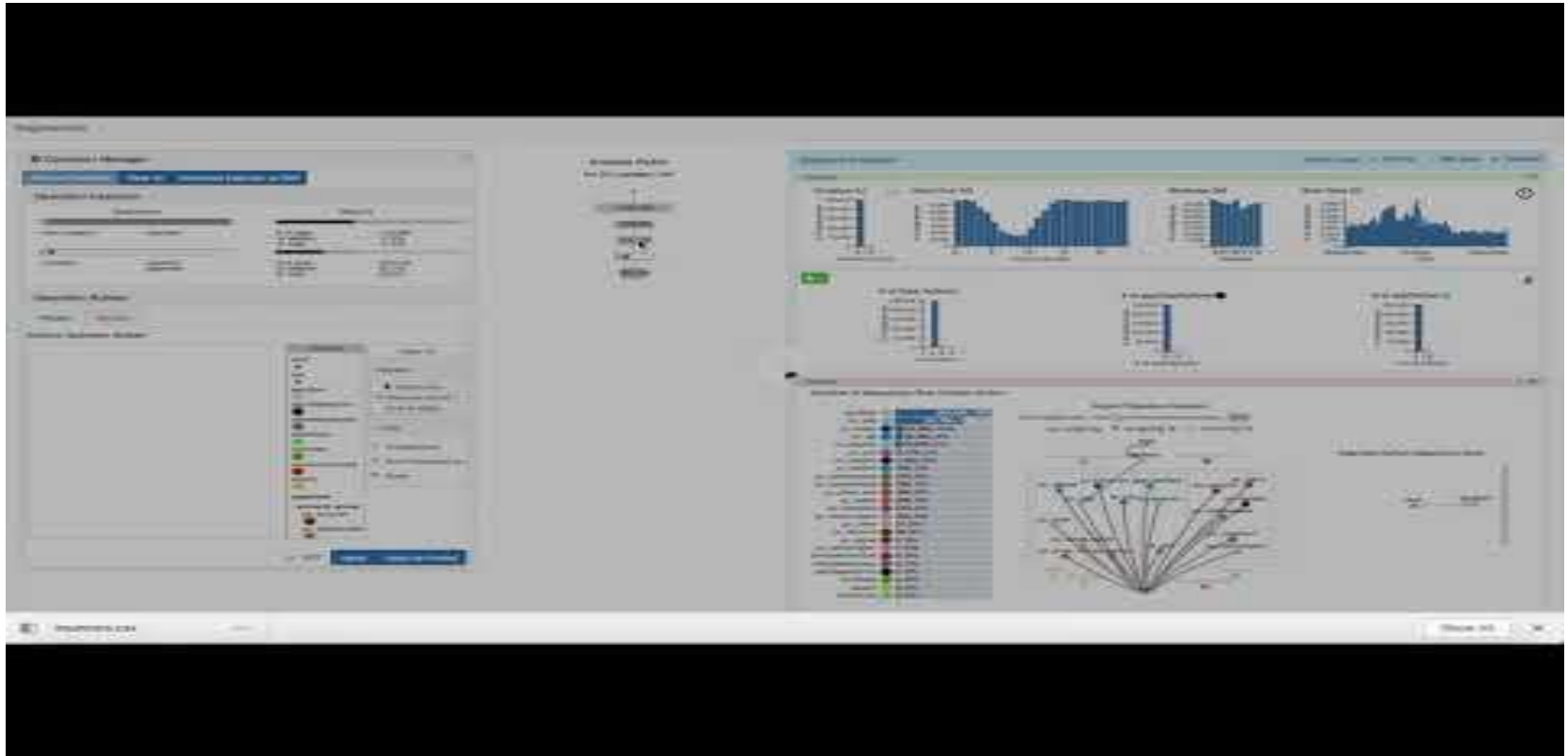


Our Solution

The Segmentifier Interface



The Segmentifier Interface

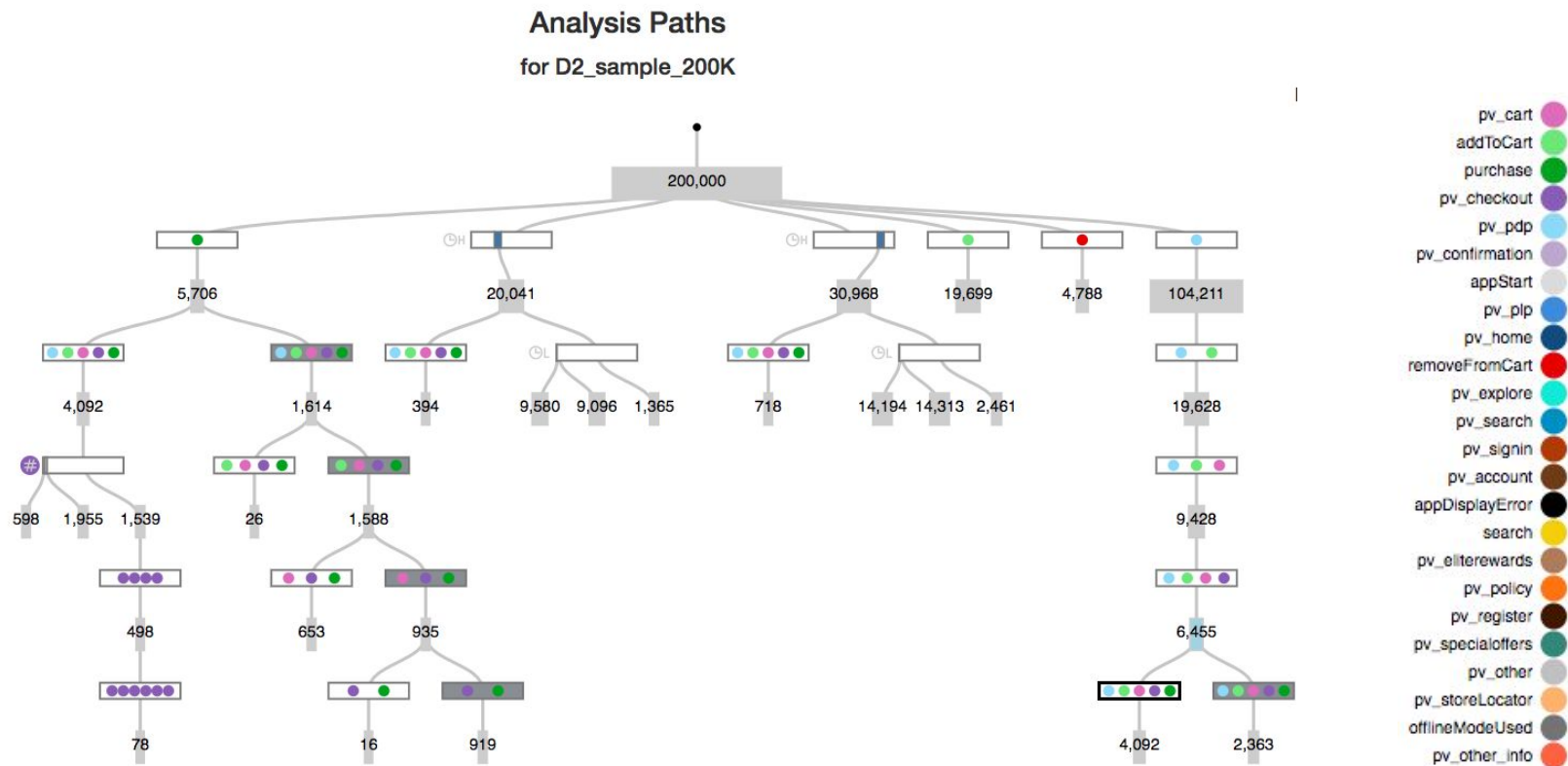


Results

Case Study #1

- 2 hour chauffeured analysis
- With industry data analyst
- Purpose:
 - One month post launch report
 - Discover actionable insights and improvements for customer
- Data
 - Session sequences
 - 200K sequences

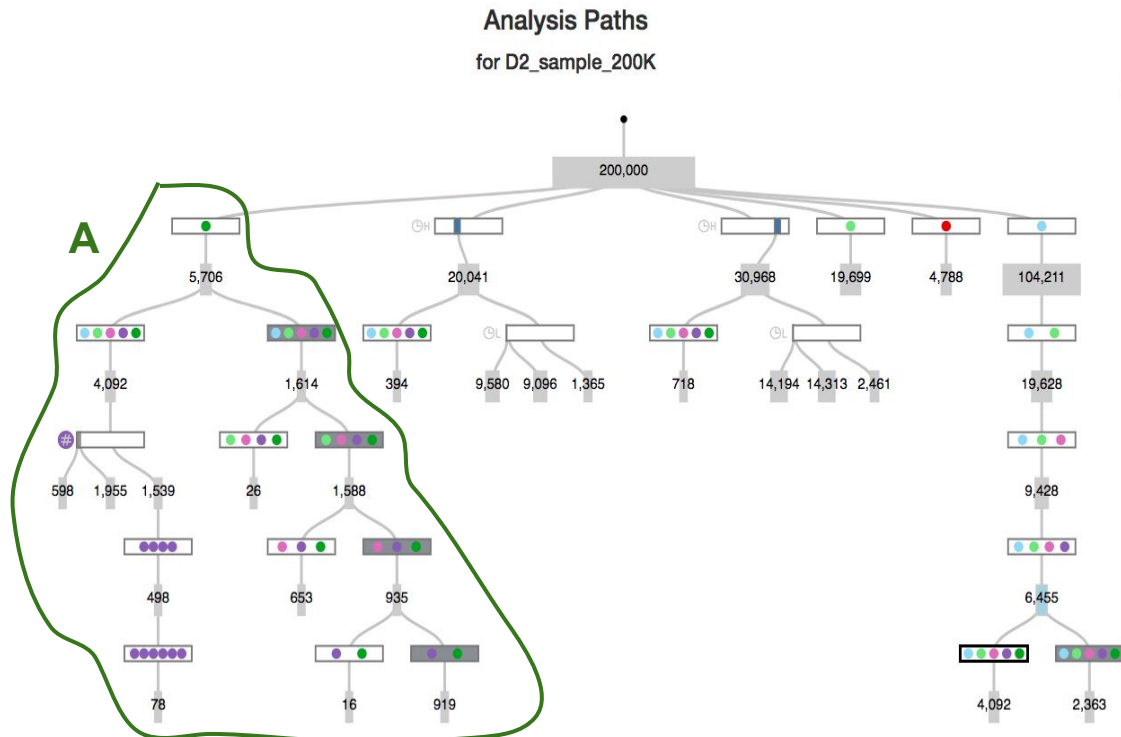
Case Study #1



Case Study #1: Analysis A

A Analyze Purchasing Behavior

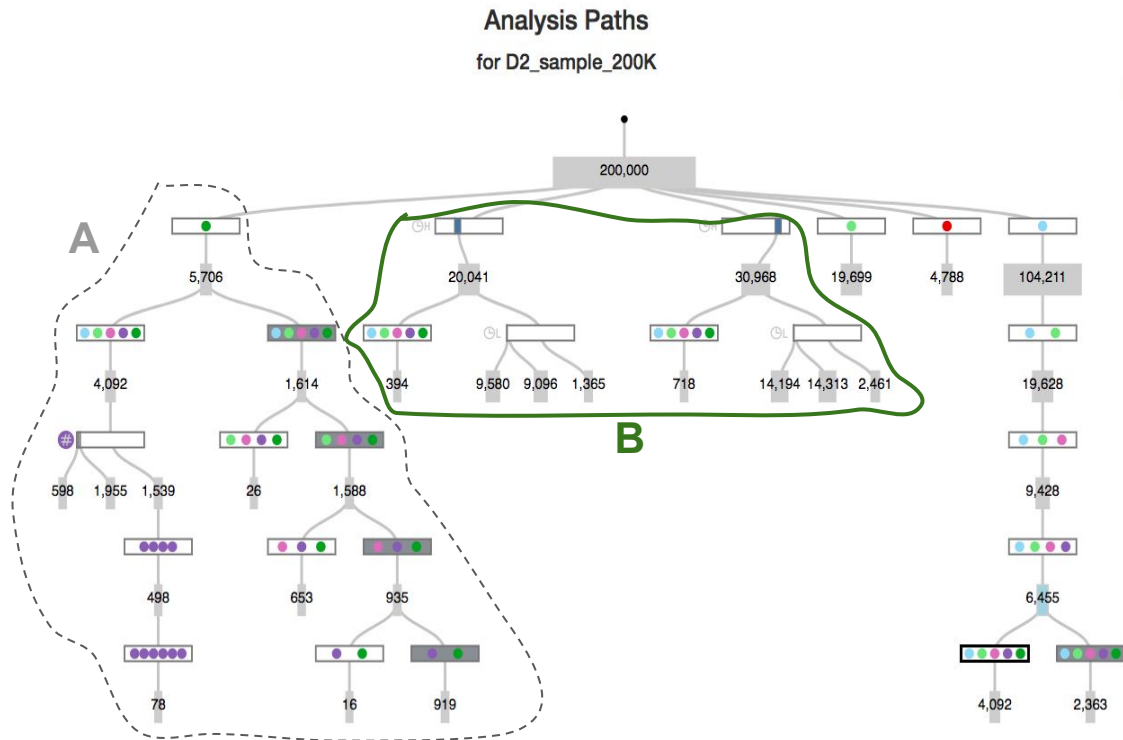
- 12% of sessions contain more checkout pages than necessary
- 30% of users actually exit the site and return later to complete their purchase



Case Study #1: Analysis B

B Compare Morning vs Night

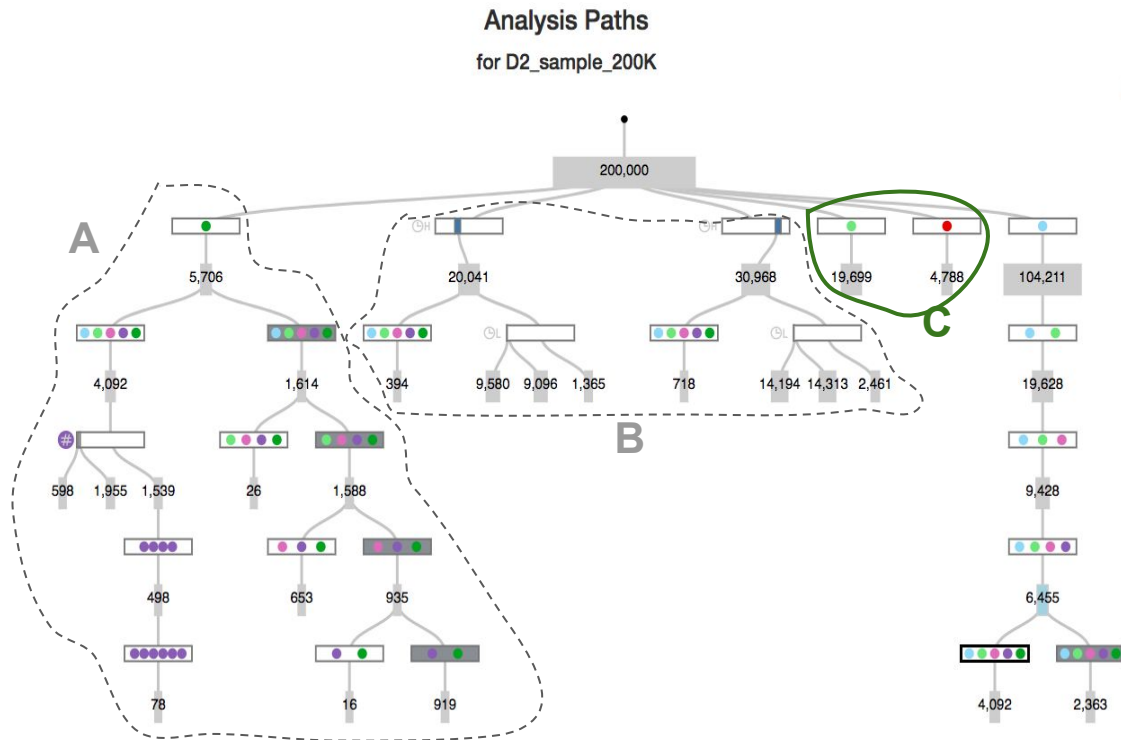
- No significant difference for percentage of sessions that contain full purchasing funnel
- No significant difference for number of actions



Case Study #1: Analysis C

C Analyze add and remove from cart behavior

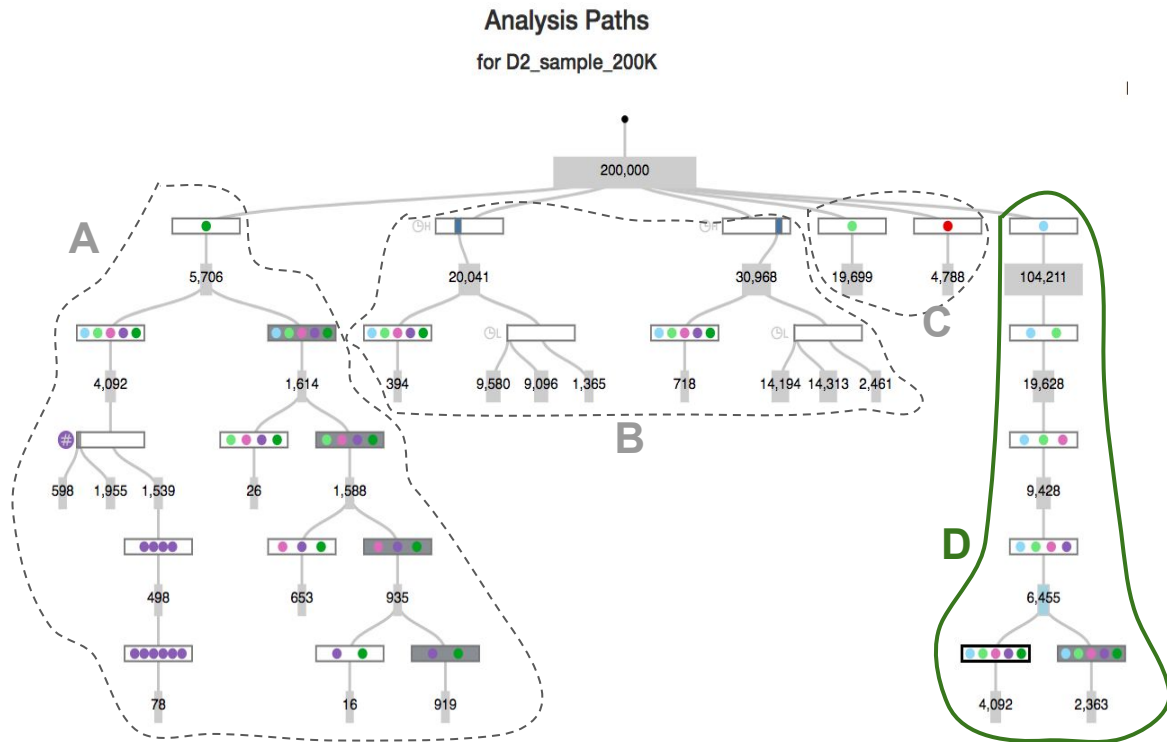
- No insight for add to cart behavior
- 30% of users who removed from cart exited the session and most likely did not come back



Case Study #1: Analysis D

D Analyze purchasing funnel

- 20% of people who get to checkout will not end up purchasing



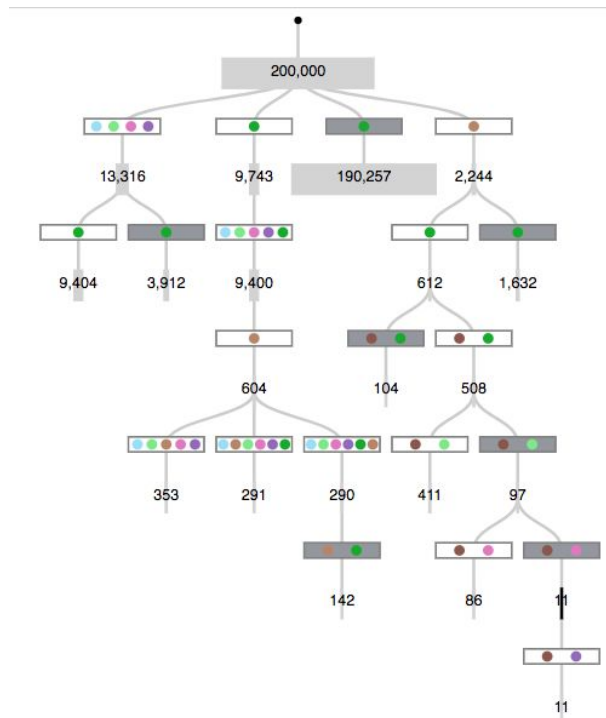
Case Study #2

- 2 hour chauffeured analysis
- With industry data analyst
- Purpose:
 - Revisit some questions from last analysis using **client sequences**
- Data
 - **Client sequences**
 - Much longer
 - Capture longitudinal behavior
 - 200K sequences

Case Study #2

Summary of Insights

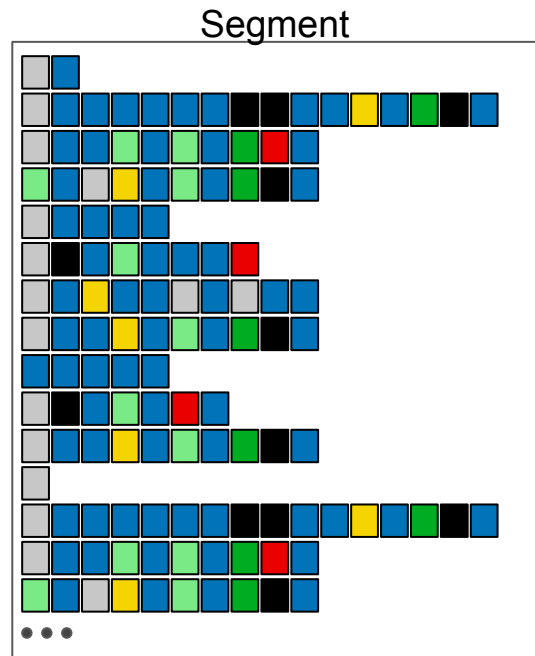
- 25% who remove from cart at checkout stage, exit and never purchase
- appStart action triggered before cart page
- Awards page analysis:
 - 1% signed up
 - 27% purchased
 - Longer sequences



Discussion

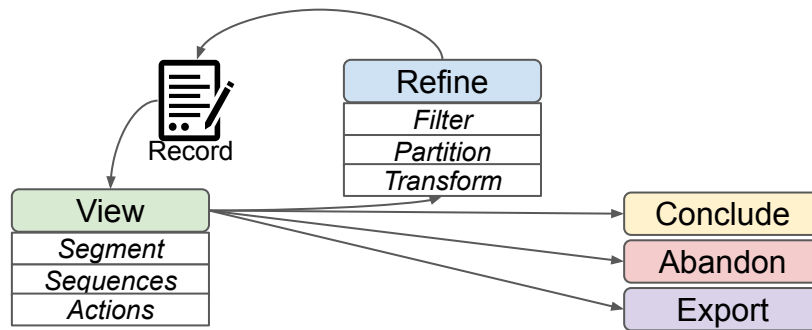
Discussion

- Goal is **Scalability**
 - Initial iterative visual refinement of large segments into useful ones
 - Attributes that align with analyst's intuitions about interesting behavior
 - Quick forming and testing of hypotheses
 - **Result:** more effective fine-grained downstream analysis



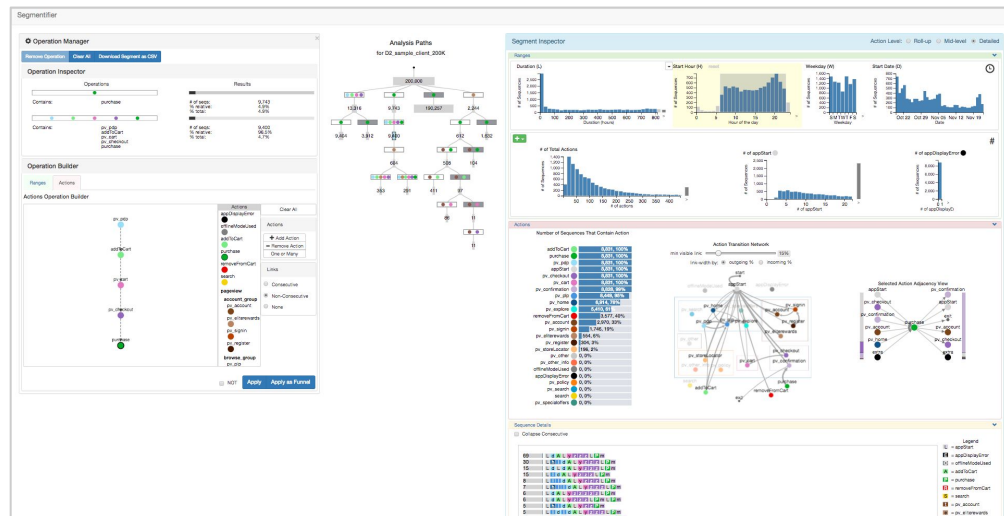
Conclusions

- Thorough **characterization of task and data abstraction** for clickstream data analysis



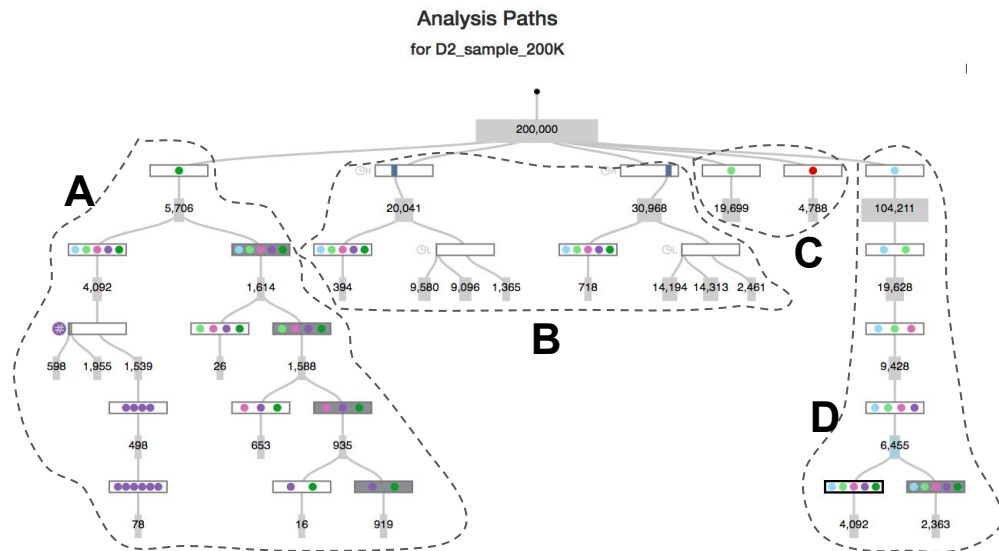
Conclusions

- Thorough **characterization of task and data abstraction** for clickstream data analysis
- **Segmentifier: novel analytics interface** for refining data segments and viewing characteristics before downstream fine-grained analysis



Conclusions

- Thorough **characterization of task and data abstraction** for clickstream data analysis
- **Segmentifier: novel analytics interface** for refining data segments and viewing characteristics before downstream fine-grained analysis
- Preliminary **evidence of utility**

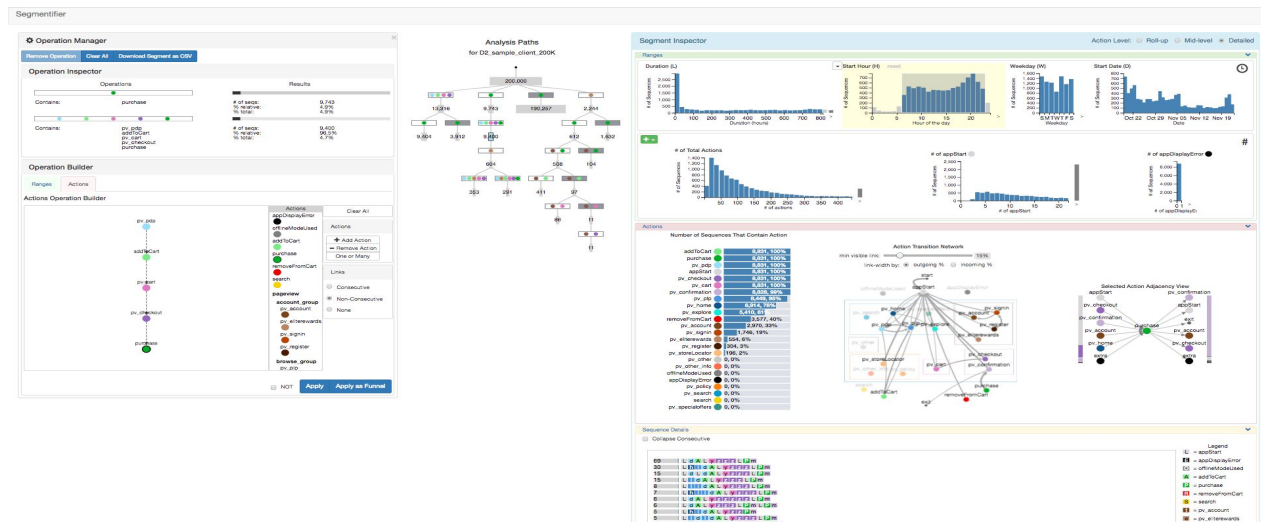


Segmentifier: Interactive Refinement of Clickstream Data

Kimberly Dextras-Romagnino, k.dextras.romagnino@gmail.com

Tamara Munzner, tmm@cs.ubc.ca, @tamaramunzner

More info: <http://www.cs.ubc.ca/labs/imager/tr/2019/segmentifier/>



Affiliations



Special Thanks



Segmentifier: Interactive Refinement of Clickstream Data

Kimberly Dextras-Romagnino, k.dextras.romagnino@gmail.com

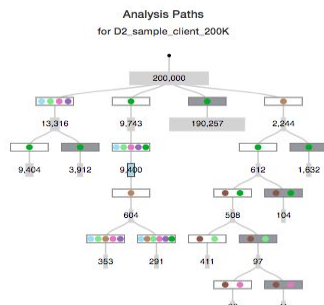
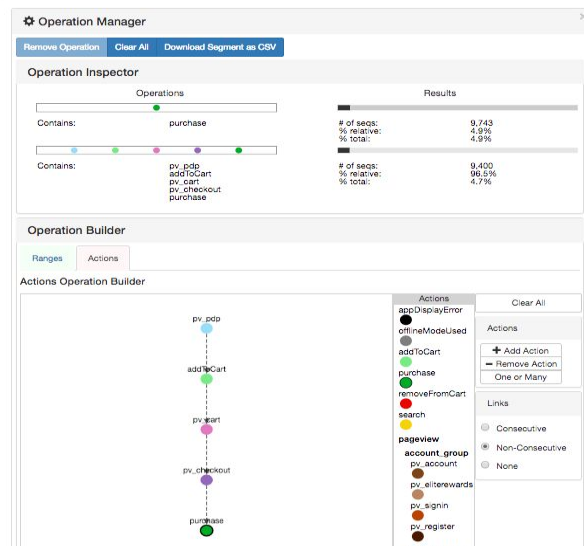
Tamara Munzner, tmm@cs.ubc.ca, [@tamaramunzner](https://twitter.com/tamaramunzner)



More info: <http://www.cs.ubc.ca/labs/imager/tr/2019/segmentifier/>



Segmentifier

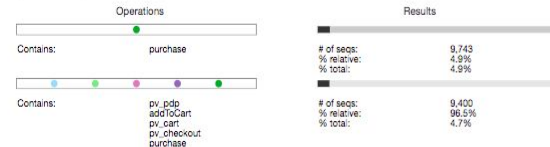


Backup Slides

⚙️ **Operation Manager**

Remove Operation Clear All Download Segment as CSV

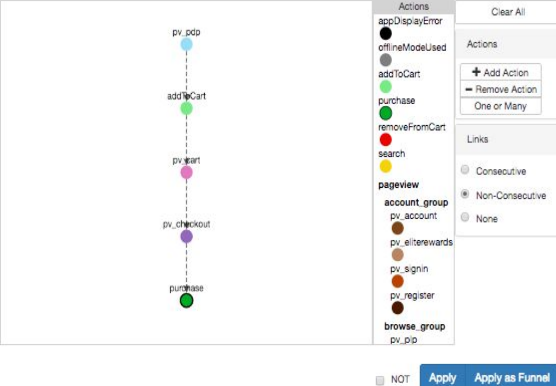
Operation Inspector



Operation Builder

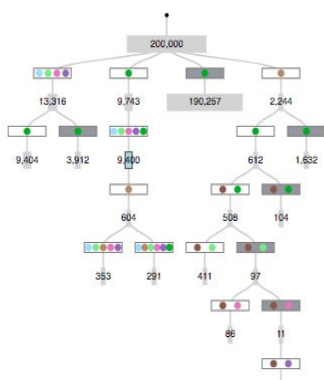
Ranges	Actions
--------	---------

Actions Operation Builder



Analysis Paths

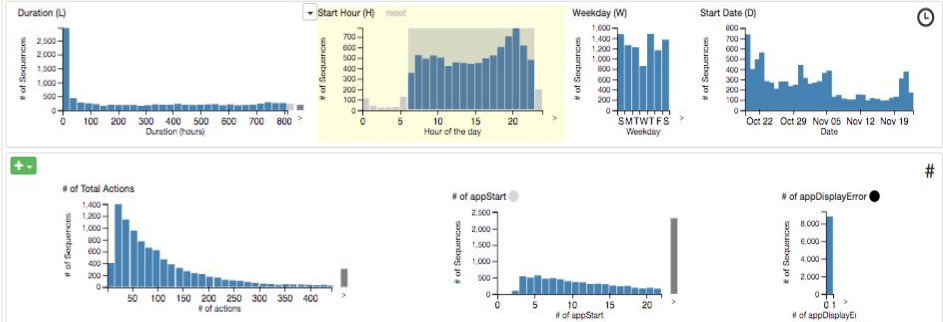
for D2_sample_client_200K



Segment Inspector

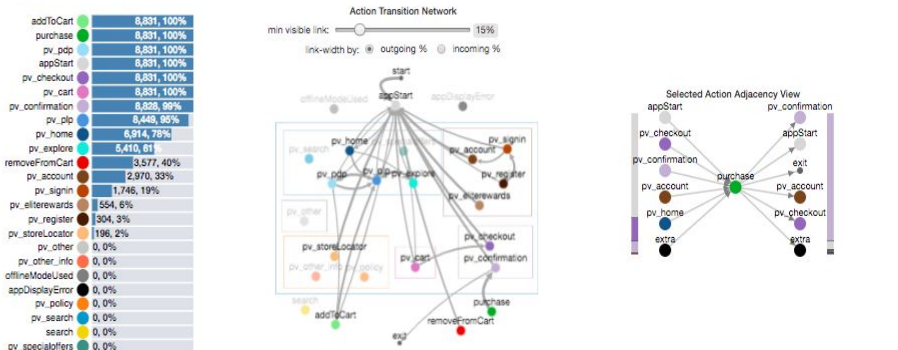
Action Level: ☐ Roll-up ☐ Mid-level ☒ Detailed

Ranges

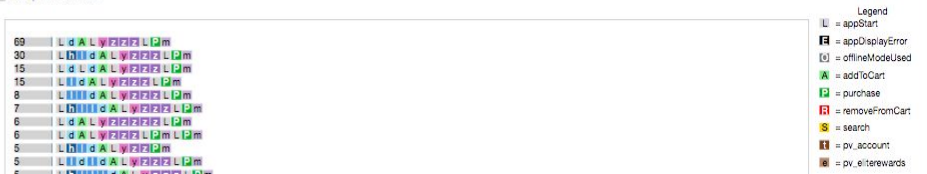


Actions

Number of Sequences That Contain Action

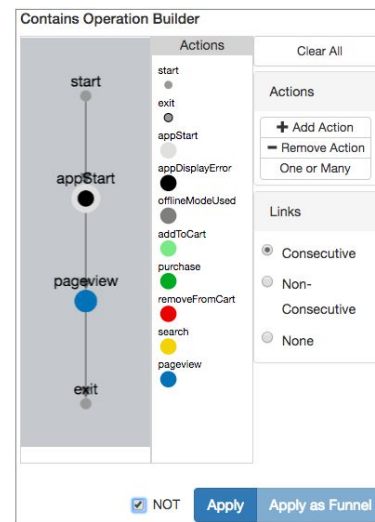
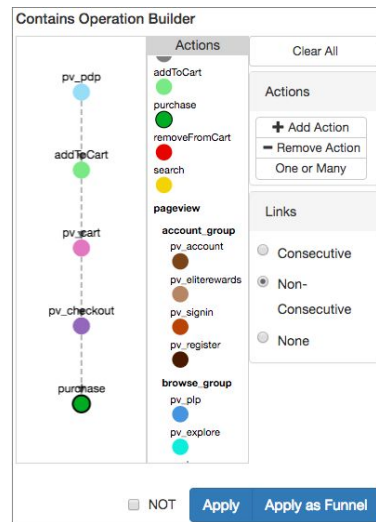


Sequence Details

☐ Collapse Consecutive

Discussion

- Trade-off between power and simplicity for this application context
 - Actions Operation Builder: regex with glyphs
 - Previous work:
 - Full support of regex
 - Difficult for non-programmers
 - Our design: deliberately less powerful so usable by non-technical analysts



Tasks: Actionable Results

Actionable Result: result or insight found through analysis that can be acted on

Result ⇒ Action

Actionable Results

Identify successful trends ⇒ Optimize

Identify problems ⇒ Fix/Improve

Identify groups of common behavior ⇒ Personalize experience

Identify site metrics/benchmarks ⇒ Keep track of state of website

Tasks: Actionable Results

Actionable Result: result or insight found through analysis that can be acted on

Result \Rightarrow Action

Actionable Results

Identify successful trends \Rightarrow Optimize

Identify problems \Rightarrow Fix/Improve

Identify groups of common behavior \Rightarrow Personalize experience

Identify site metrics/benchmarks \Rightarrow Keep track of state of website



Domain-Specific Questions

How many users purchase? What path did they choose?

Tasks: Actionable Results

Actionable Result: result or insight found through analysis that can be acted on

Result \Rightarrow Action

Actionable Results

Identify successful trends \Rightarrow Optimize

Identify problems \Rightarrow Fix/Improve

Identify groups of common behavior \Rightarrow Personalize experience

Identify site metrics/benchmarks \Rightarrow Keep track of state of website



Domain-Specific Questions

How many bounce (exit after viewing one page)?

Tasks: Actionable Results

Actionable Result: result or insight found through analysis that can be acted on

Result ⇒ Action

Actionable Results

Identify successful trends ⇒ Optimize

Identify problems ⇒ Fix/Improve

Identify groups of common behavior ⇒ Personalize experience

Identify site metrics/benchmarks ⇒ Keep track of state of website



Domain-Specific Questions

Can you classify different types of buying behaviors?

Tasks: Actionable Results

Actionable Result: result or insight found through analysis that can be acted on

Result \Rightarrow Action

Actionable Results

Identify successful trends \Rightarrow Optimize

Identify problems \Rightarrow Fix/Improve

Identify groups of common behavior \Rightarrow Personalize experience

Identify site metrics/benchmarks \Rightarrow Keep track of state of website

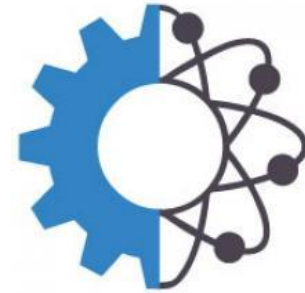


Domain-Specific Questions

What is the average number of sessions in a month? Was this month abnormal?

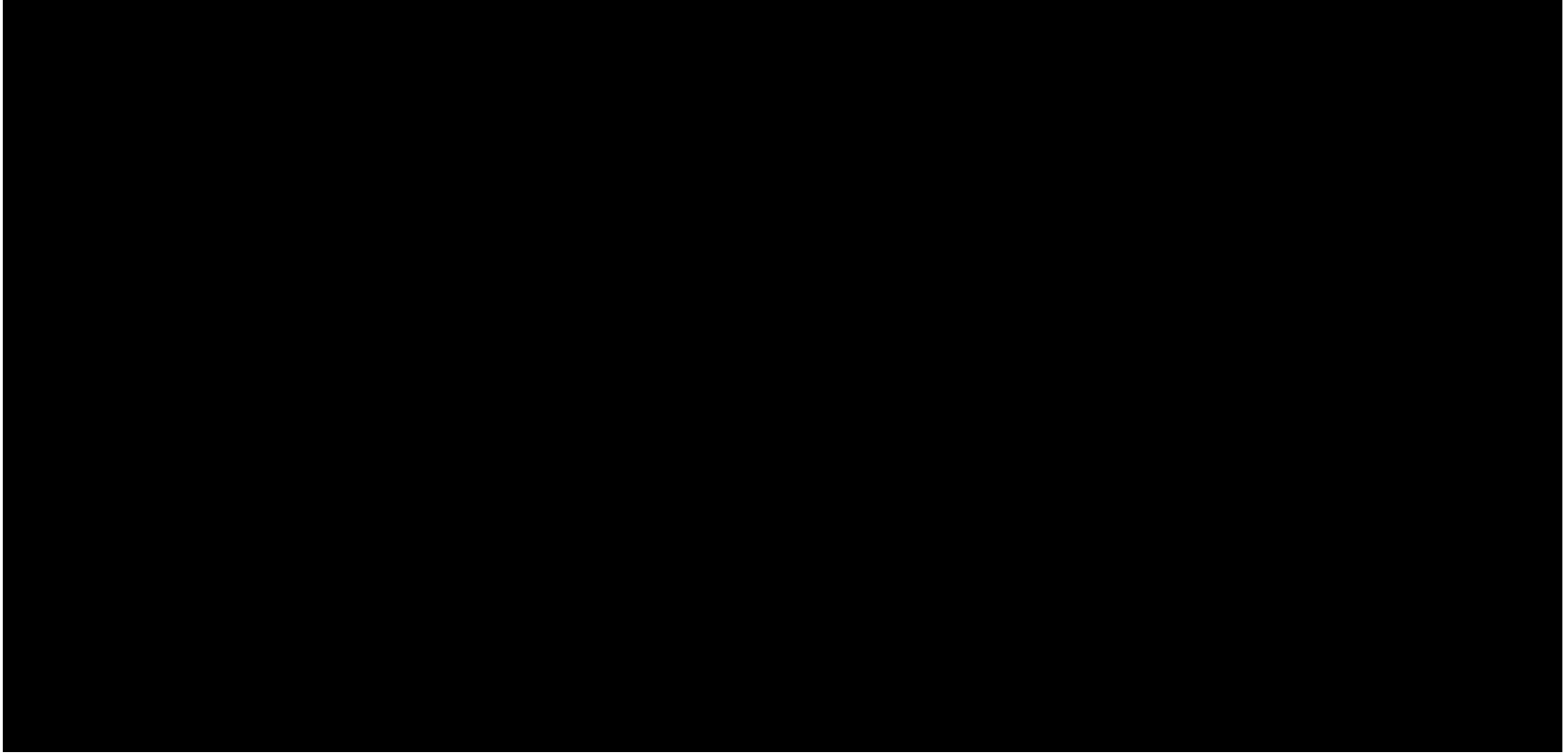
Discussion + Future Work

- Focus on agile and iterative development of design
 - Modest engineering effort to achieve base level of usability to test design concept
 - Loading times
 - Processing time
 - Goal:
 - Proof of concept that design works for target tasks
 - Not (premature) engineering optimization
 - Future work:
 - Engineering optimization for this final design



Extra Slides

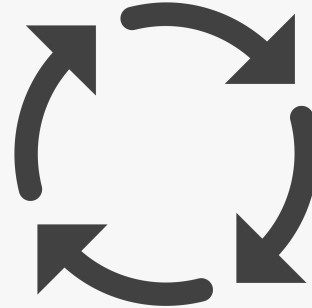
The Segmentifier Interface



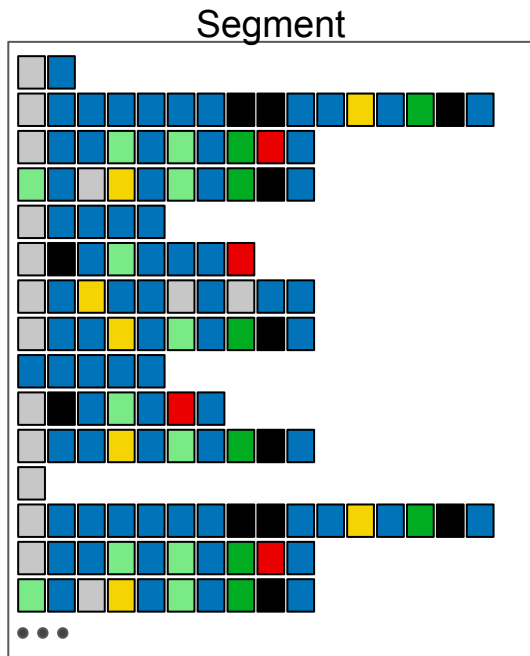
<https://www.youtube.com/watch?v=TobYDFeISOg&t=24s>

Research Method: Mobify

- **Pre-condition Phase**
 - Period of 5 months
 - Met with 12 employees
- **Core Phase**
 - Data and Task Abstraction
 - Design interface
 - Implement interface
- **Analysis Phase**
 - Formulate Framework
 - Write Paper/Thesis

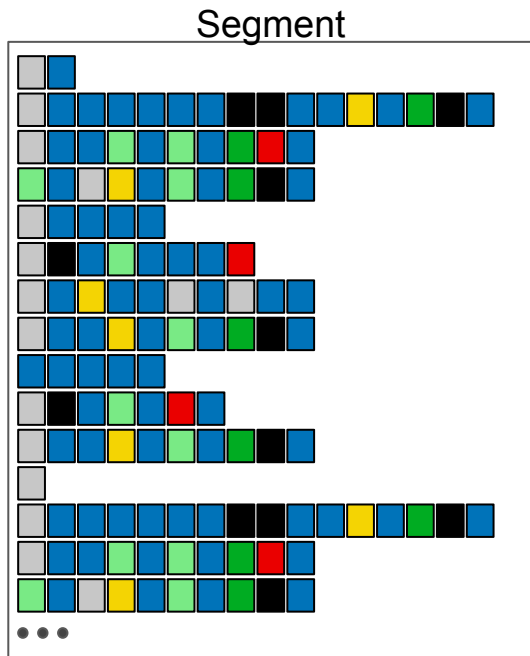


Real-world Clickstream Data



Scale is huge

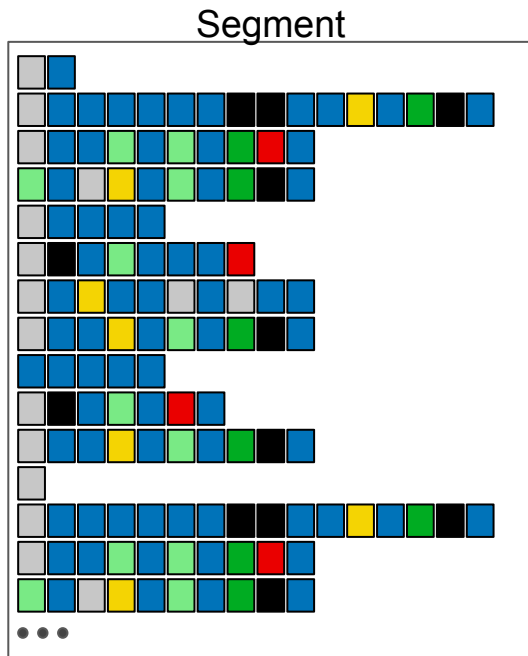
Real-world Clickstream Data



Scale is huge

Variability is high

Real-world Clickstream Data



Scale is huge

Variability is high

Most work **fails** when applied to real-world data.

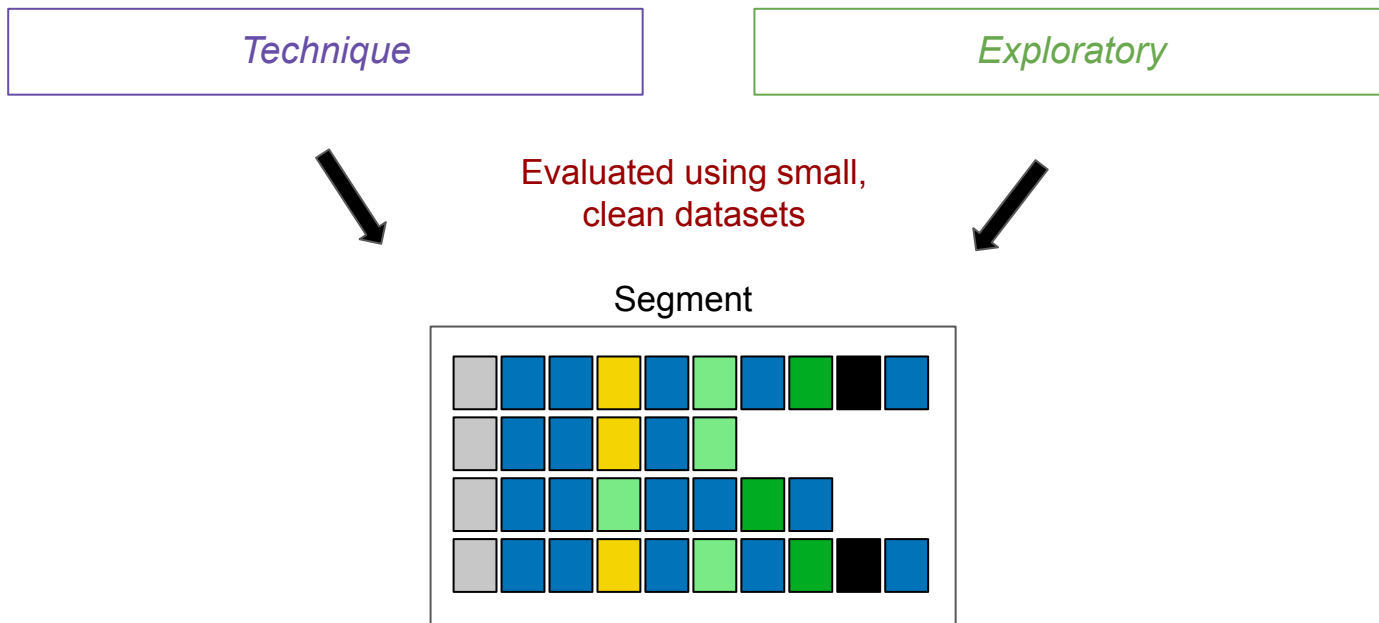
Technique

Most techniques have data requirements to work effectively

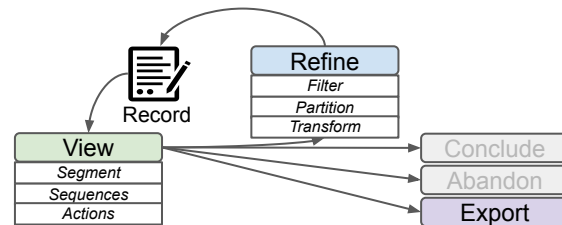
Exploratory

Most focus on analyzing sequences.
Too many to view at once.

Related Work: Problems



Related Work



Export

Post-Export: Specific Techniques

- Clustering [Wei et al.], Pattern Mining (CoreFlow [Liu et al.], Frequence [Perer et al.]
- Require small, clean datasets

View

View Sequences: Event Sequence Visual Overviews

- CareFlow [Perer et al.]
- Can't refine segments or view segment attributes

Refine

Refine: Visual Query Systems

- i.e. COQUITO [KPS16], (s|qu)eries [ZDFD15], DecisionFlow [GS14], PatternFinder [FKSS06], and SparqlFilterFlow [HLBE14]



Record

Record: Graphical Histories

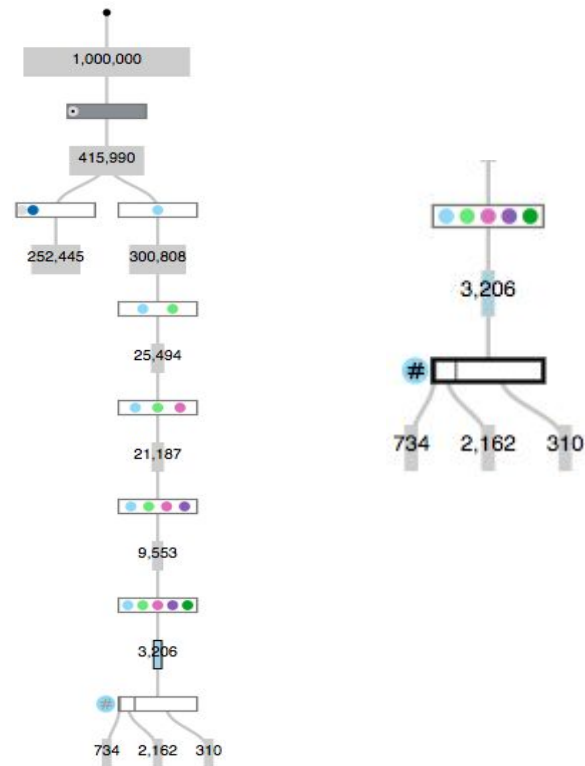
i.e.

Tasks: Task Abstraction

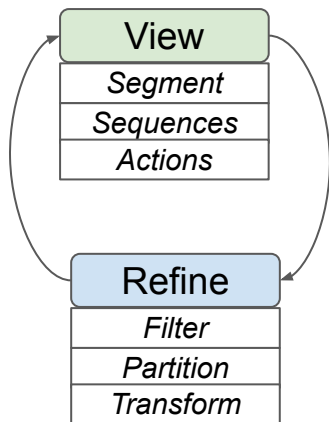
- **T1) Identify:** Find some set of sequences that constitutes interesting behavior
 - consumers in loyalty program browse longer
- **T2) Drilldown:** Distinguish more specific behaviors to further partition a segment previously defined by looser constraints
 - check if purchasers fall into natural groups by time of day
- **T3) Frequency:** Determine how many sequences are in the segment defined by behavior X
 - check ratio of bouncers to non- bouncers
- **T4) Ordering** within sequence: Match if action subsequence X occurs before (or after) action subsequence Y in a sequence
 - verify that all users add to cart before purchasing

Discussion + Future Work

- Understandable segments:
 - Each possible refinement operation corresponds to one attribute constraint
 - In contrast to clustering, pattern mining that have uninterpretable results for this scale of noisy data
- Segmentifier explicitly supports refinement through both filtering and partitioning.
 - Encourages subsequent analysis
 - Allows comparison
 - Future comparison work



Related Work



View and Refine: Filtering Sequences To Segments

- SessionViewer [Lam 2007], EventFlow [Munroe 2013] , EventPad [Cappers 2018]
- Lack of segment attributes
- Lack of ability to record analysis path

