



Smithsonian
Cooper-Hewitt, National Design Museum

AK

THE TRAILS IN US

PATTERNS OF HUMAN BEHAVIOR



akamlani@gmail.com | arikamlani.com



[@akamlani/cooperhewitt](https://github.com/akamlani/cooperhewitt)

THE PEN ARRIVES

WHY THE PEN



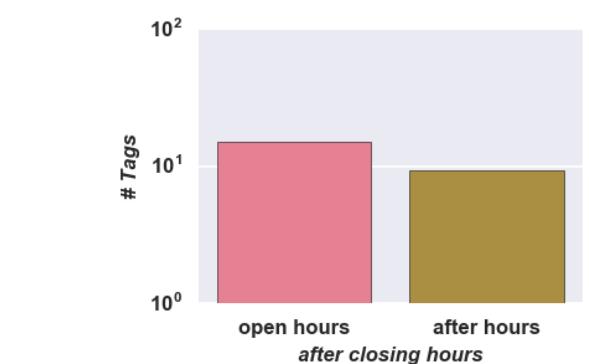
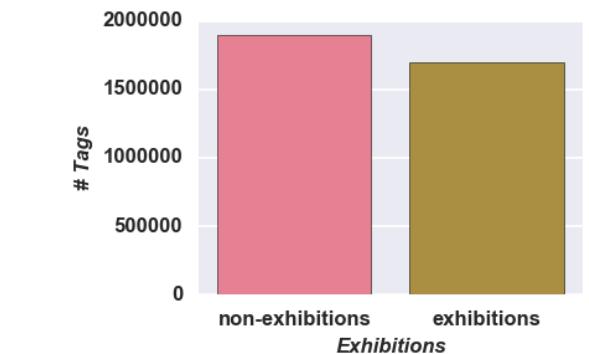
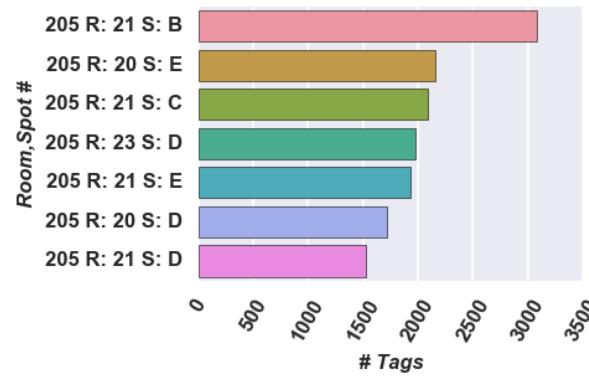
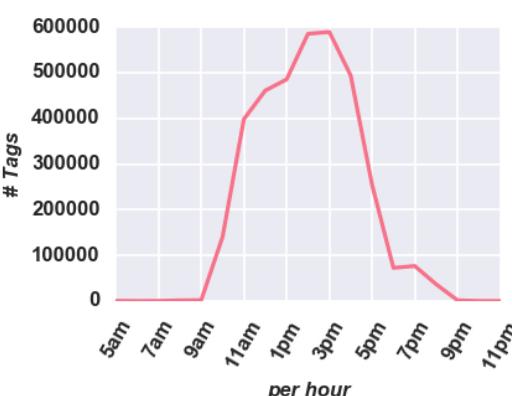
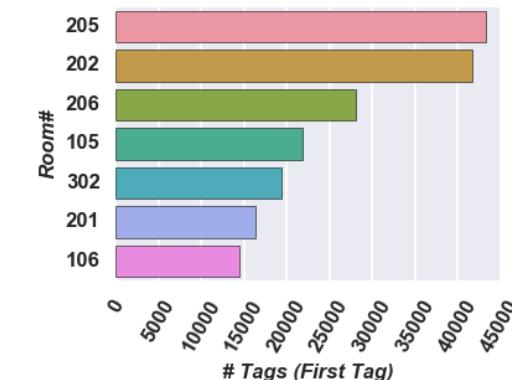
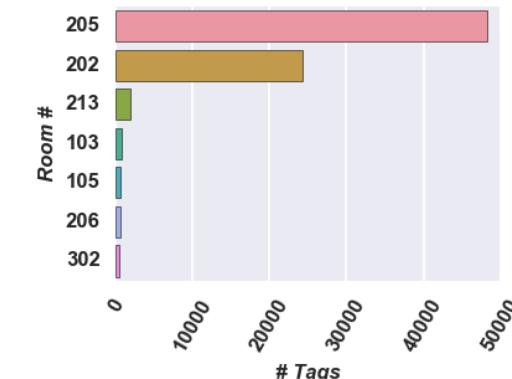
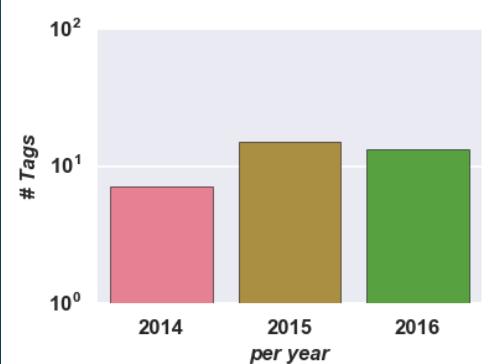
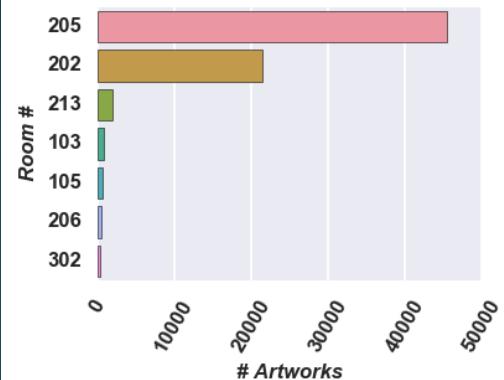
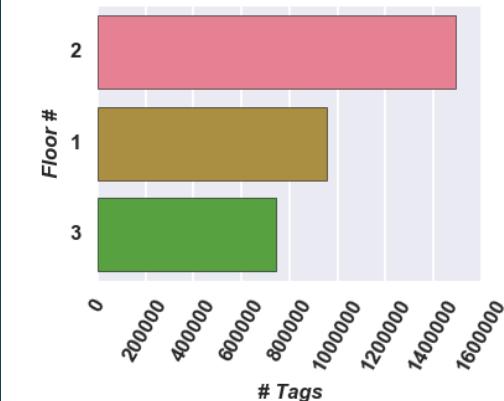
AK

- ▶ ENHANCE VISITOR EXPERIENCE
 - ▶ DRAW VISITOR INTO THE DESIGN EXPERIENCE
- ▶ RECENT DATASET RELEASED MAR 2016
 - ▶ 3.5M+ OBSERVATIONS
 - ▶ AGGREGATE W/ COLLECTION METADATA (80K+)
- ▶ PROJECT FOCUS:
 - ▶ VISITOR SEQUENCE PATTERN TAGGING BEHAVIOR
 - ▶ IDENTIFYING SEGMENTS BASED ON VISITORS TAG BEHAVIOR



LOVE THY DATA

AK



- ▶ HIGH CAPACITY ROOMS
 - ▶ RM 205,202
- ▶ FIRST ARRIVAL
 - ▶ RM 205, 206, 202
- ▶ EXHIBITIONS
 - ▶ INFERRED FROM CREATED DAILY BUNDLES
- ▶ AFTER HOURS
 - ▶ MUSEUM TESTING DATA
 - ▶ SPECIAL EVENTS

PAGE RANK



TOP 10 INFLUENTIAL ARTWORKS BASED ON A VISITOR'S DAILY JOURNEY
MODELING THE SEQUENCES



RM 206



RM 212/105



RM 201



RM 212/105



RM 201



RM 212/105



RM 302



RM 206



RM 105



RM 302

GROUP A DAILY JOURNEY AS AN UNWEIGHTED DIRECTED SEQUENCE

► THEMES & DEPARTMENTS

- STAIRCASES
- MURALS, POSTERS
- TEXTILES

► TRANSITION SPOTS

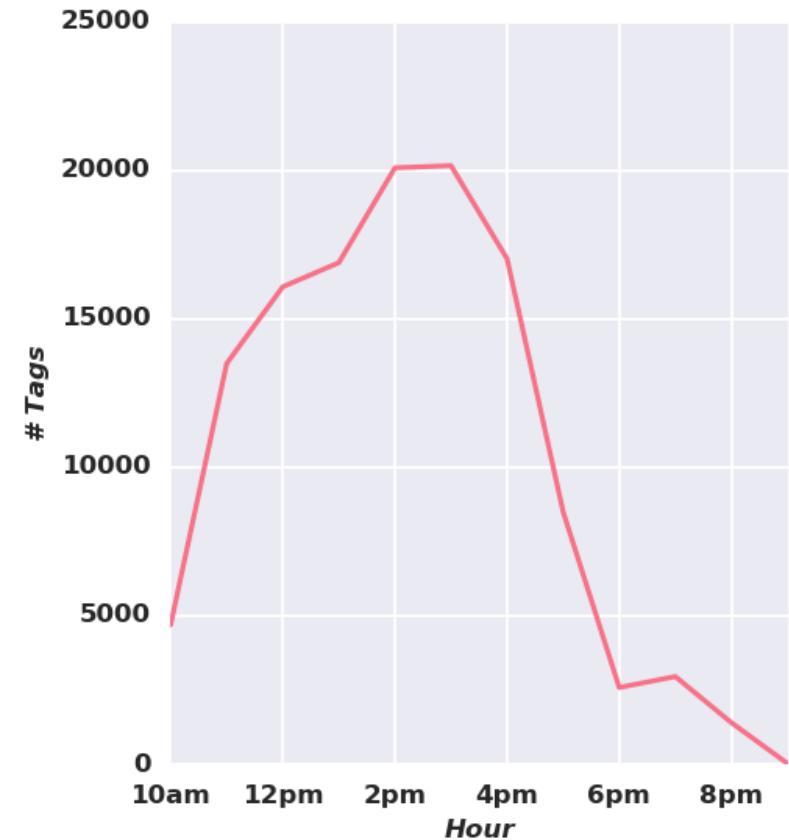
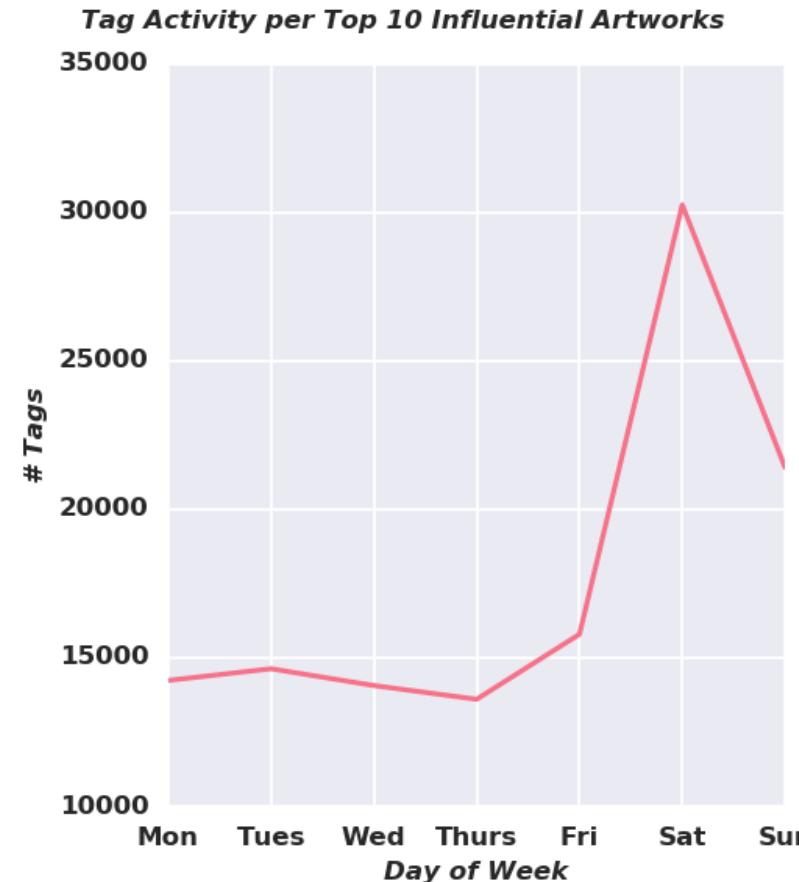
- POSTERS (RM 202, 205)
- DRAWINGS (RM 202, 205)
- CONCEPT ART (RM 103)

► % OVERALL TAGS: ~ 3%

► % EXHIBITS: 30%

PERIODS OF ACTIVITY

Per Top 10 Influential Artworks



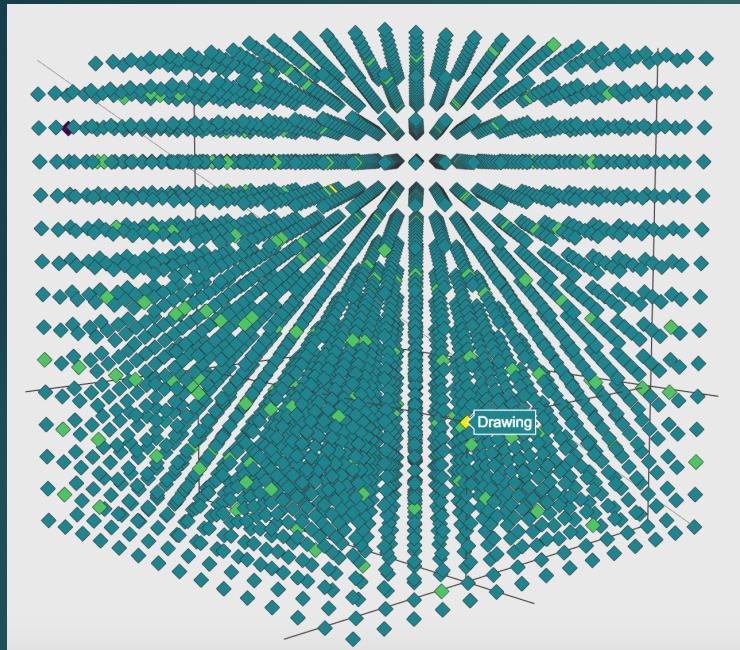
SEASONAL PEAKS: SUMMER, WINTER

WEEKEND PEAKS

AFTERNOON PEAKS: 2-4 PM

COMMUNITIES

LABEL PROPAGATION ALGORITHM (LPA)



SAMPLE OF COMMUNITY DETECTION



- ▶ ONE LARGE DENSE COMMUNITY
- ▶ ONE SMALL COMMUNITY
- ▶ SEVERAL SMALL MICRO-COMMUNITIES

DENSE COMMUNITY	SMALL COMMUNITY
EVEN DISTRIBUTION OF EXHIBITIONS	N/A
HIGH CAPACITY ROOMS (RM 205,202) THEME ROOMS (1 st FLOOR: PROCESS LAB)	SMALL CAPACITY ROOMS SPECIFIC TO 3 RD FLOOR (302)
TYPES: POSTERS/DRAWINGS/CONCEPT ART	TYPES: VESSELS/SKETCHES/PRINTS
DEPT: DRAWINGS/PRODUCT DESIGN/...	DEPT: PRODUCT DESIGN
AFTER HOURS SPECIAL EVENTS	N/A

WE CAN DO BETTER...

TEMPORAL PATTERNS



MORE THAN SEQUENCES

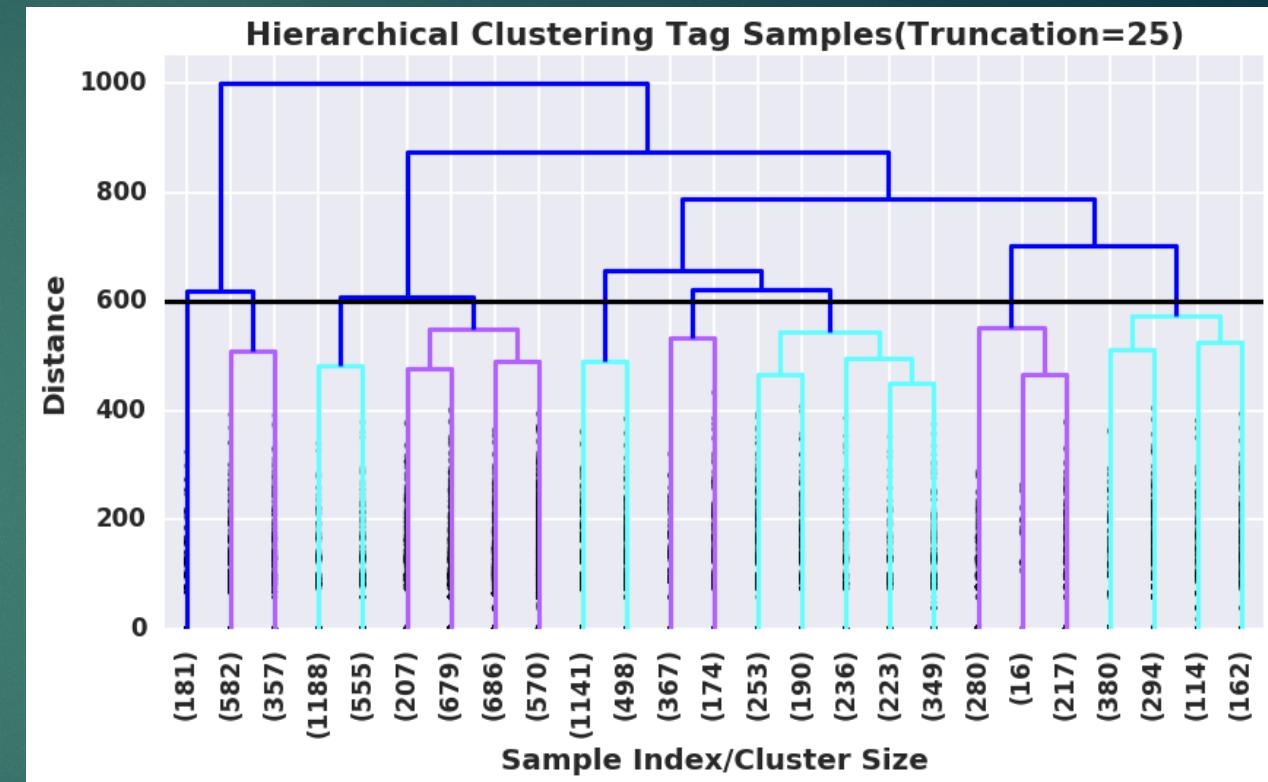
- ▶ BURSTY VISITOR SEGMENTS
 - ▶ RESAMPLED TIME PERIOD SEQUENCES
 - ▶ AS OUTLIERS BASED ON MOVING AVERAGE (MA) WINDOW
- ▶ ACTIVITY TIME PERIODS
 - ▶ PEAK HOURS, DAYS OF WEEK, TIME OF YEAR
- ▶ ARTWORK LOCATIONS
 - ▶ ARTWORKS WITH MULTIPLE LOCATIONS BASED ON TIME PERIOD
 - ▶ SPOT LOCATIONS WITH A LARGE TURNOVERS
 - ▶ CAPACITY BASED ROOMS
- ▶ TAGS ASSOCIATED TO EXHIBITIONS

FEATURE ENGINEERING

HIERARCHICAL CLUSTERING

BASED ON TEMPORAL PATTERNS

- ▶ ALL DATA IS CATEGORICAL/BINARY
 - ▶ KMEANS NOT A GOOD FIT
 - ▶ HIERARCHICAL CLUSTERING (AGGLOMERATIVE)
 - ▶ COPH DISTANCE (CITYBLOCK) ~ 0.73
 - ▶ COMPUTATIONALLY EXPENSIVE (SAMPLE)
(NOT AVAILABLE IN SPARK MLlib/ML)
- ▶ IDENTIFIED CLUSTERS
 - ▶ FLOORS
 - ▶ ARTWORK CAPACITY BASED ROOMS
 - ▶ HIGH VS CONSTANT SPOT ARTWORK TURNOVER
 - ▶ INTERACTIVE VISITOR STATION AREAS
 - ▶ SPECIAL EVENTS (AFTER HOURS)



RECOMMENDATIONS

HOW TO HELP THE MUSEUM

- ▶ FURTHER IMPROVE ON VISITOR EXPERIENCE – NO MISSED OPPORTUNITIES
- ▶ OPPORTUNITY TO IMPROVE ON TRANSITION SPOT LOCATIONS
- ▶ TAILORED VISITOR SEGMENTATIONS

FUTURE VISION

- ▶ ADDITIONAL MODELS ON MEMORY SEQUENCES (RNN/LSTM)
- ▶ FEATURE ENGINEERING ON ANNOTATED METADATA (NLP)
- ▶ TRANSFER LEARNING (APPLY TO OTHER DOMAINS)

DISCUSSION

AK



Website: arikamlani.com

Email: akamlani@gmail.com



[@akamlani](https://twitter.com/akamlani)

AK

Appendix

COMPONENTS



NON-DISTRIBUTED



igraph

DISTRIBUTED

