TIME SPENT ONLINE INCREASED OVER 75% IN 4 YEARS

Time Spent Online on Key Internet Categories
Source: comScore Media Metrix, Worldwide, March 2007 - October 2011

Legend: ■ Other Content ■ Communications ■ Retail ■ Search/Navigation ■ Social Networking

comSCORE. Data Gem
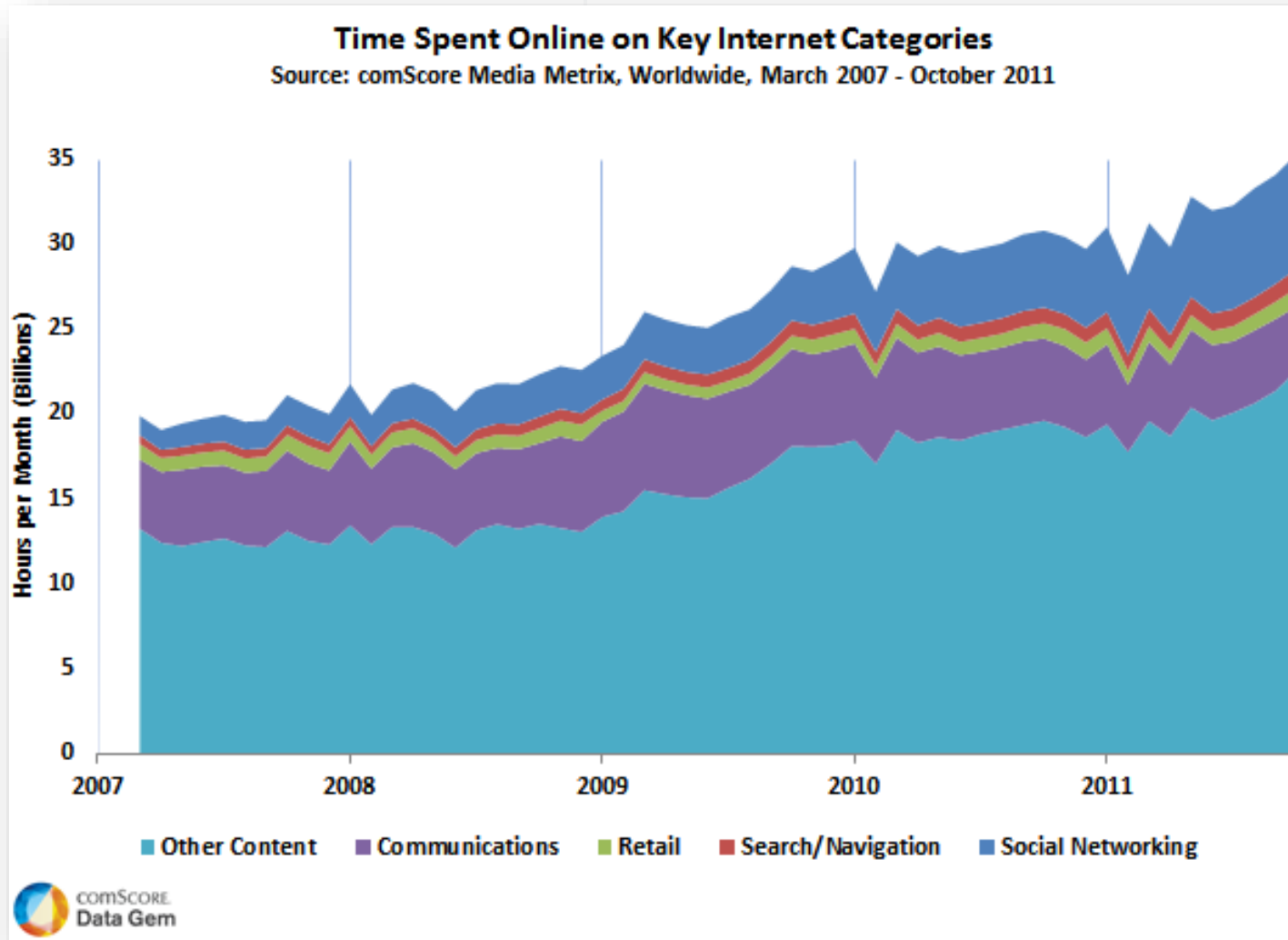
Total Time Spent Worldwide
As a percentage of total Internet hours spent by all users:

Social Networking 22%

Viewing Content 42%

Other (includes email, commerce, search) 36%

(Figures may vary slightly as categories tend to overlap)
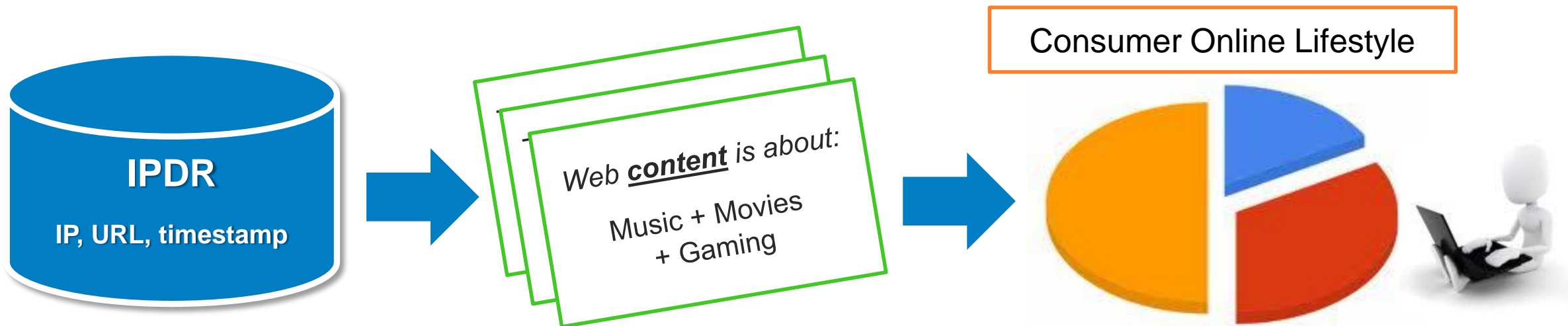sources: Pew, Nielsen

SAS | THE POWER TO KNOW.
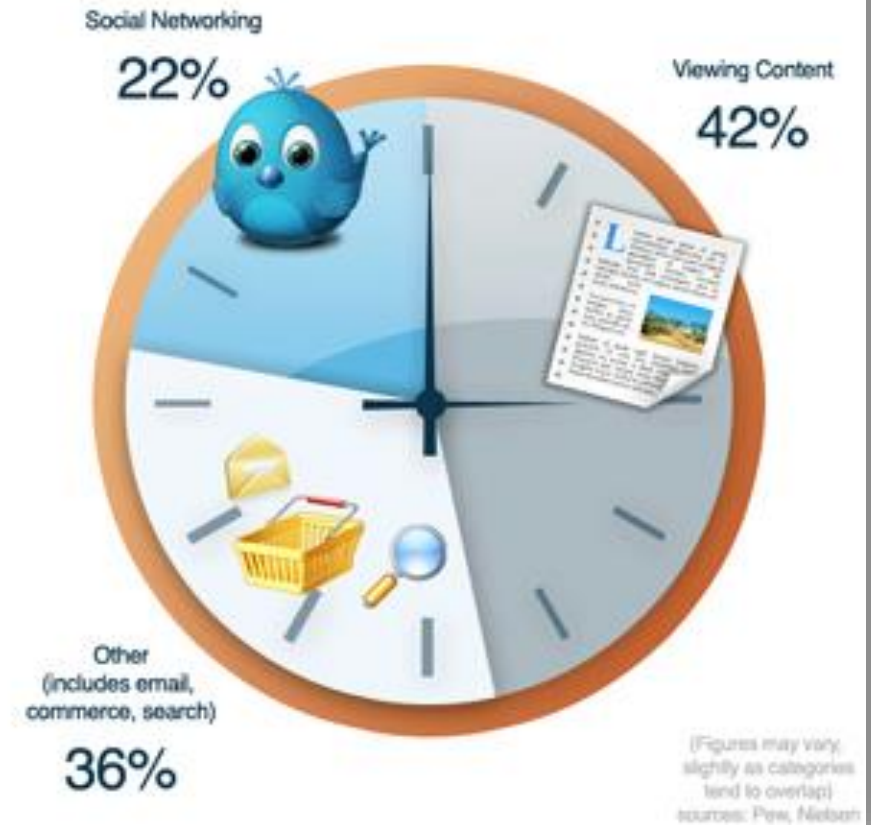
WHAT IF WE KNEW WHO IS INTERESTED IN WHAT, WHERE AND WHEN?

- If we knew facts such as –

  - **"60% of time spent only by residents of town X last month was about Rock music and Action movies"**

- We can better improve initiatives such as –

  - Contextual marketing

  - Campaign/ theme design

  - Up-sell/ cross-sell

  - Improving service quality

  - Building consumer communities

  - And more…

- Unique opportunity for CSPs to analyze IPDR data
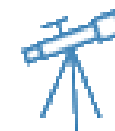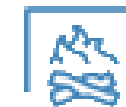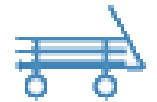
SOLUTION | HOW DO WE ANALYZE ONLINE LIFESTYLES?

- Explore consumer web surfing behavior in IP data records (IPDR)

  - IPDR describes who (IP), what (URL), when (timestamp) of web surfing behavior

- Categorize web surfing behavior via a deep rooted taxonomy

  - E.g. Top/Entertainment/Video Games/Music_video_games;

  - Based on standards in International Press Telecommunications Council - www.iptc.org

- Visualize the results over various dimensions to discover online lifestyles of consumers

  - Visualize interests against customer segments/ profile, events, location, etc

IPDR

IP, URL, timestamp

Web **content** is about:

Music + Movies + Gaming

Consumer Online Lifestyle

§sas | THE POWER TO KNOW.

CHALLENGES | HOW DO WE MAKE SENSE OF THE CONTENT?

1. The IPDR/ URL may not enough to determine the content of the URL

   - **Logs describes only Timestamp, IP, URL**

2. For example, If we look at: **www.ABC.com**

   - **How do we know if it refers to:**

     a school? or a media firm? or a beer?
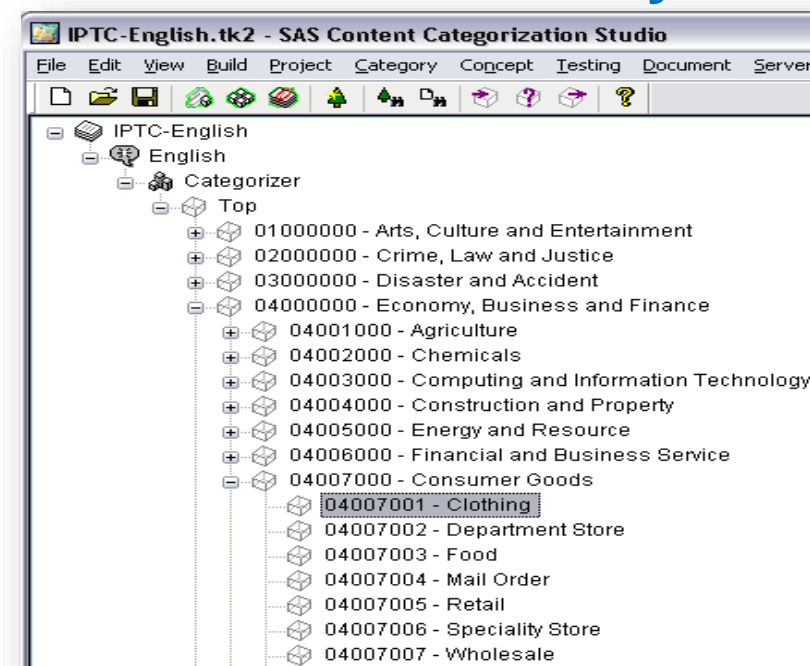

or

or

CHALLENGES | SOLUTION

1. In order to accurately categorize a URL, we will look at the content in the URL

2. Then we categorize the content based on IPTC taxonomy

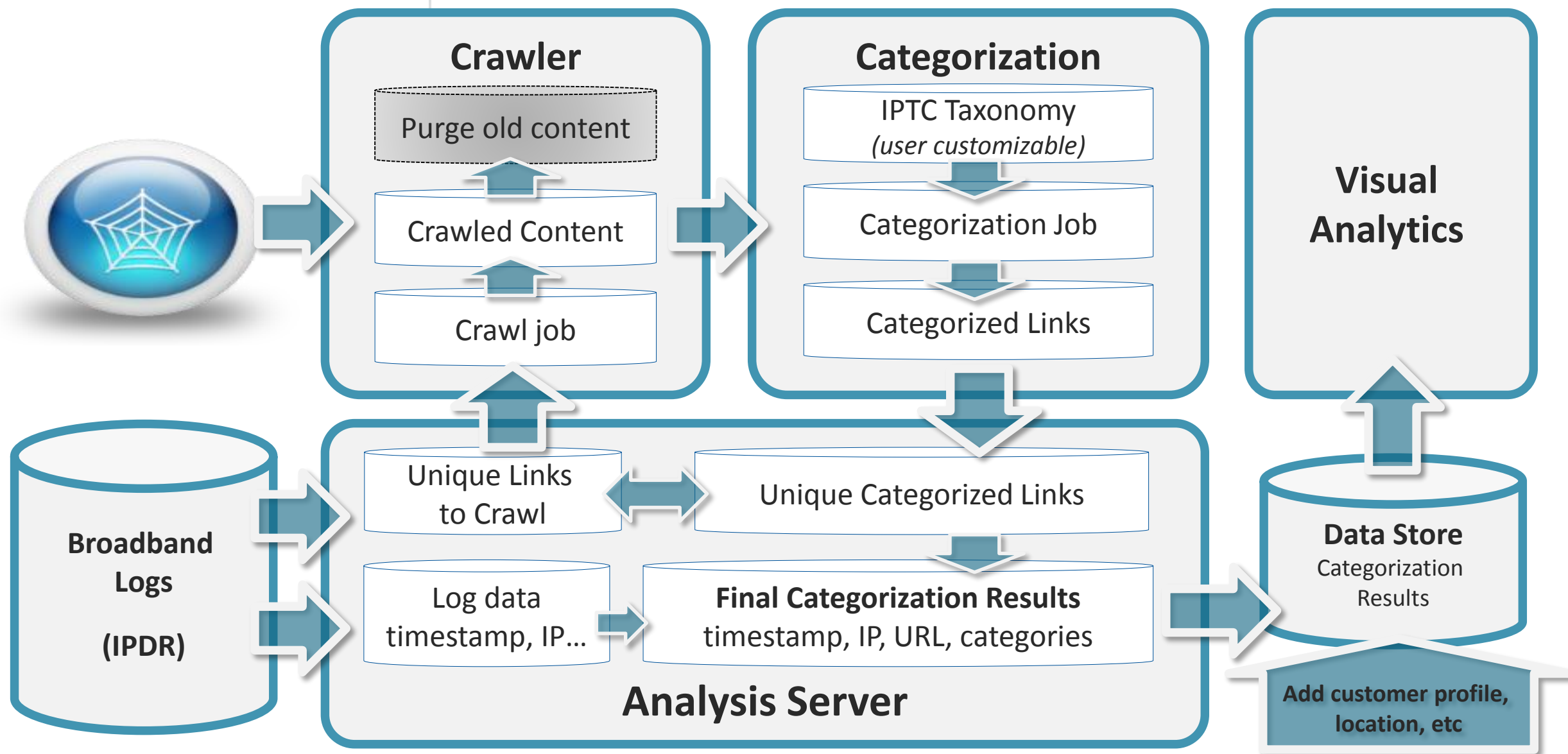3. Weekly/ monthly reports and sample results are generated to check scoring accuracy

1. Homepage of the URL

News, Topics, Reports…

Stout, beer, roasted malt…

Courses, Fees, Reading…

2. IPTC taxonomy

UNDER THE HOOD | HOW DOES IT WORK?

Crawler

- Purge old content
- Crawled Content
- Crawl job

Categorization

- IPTC Taxonomy *(user customizable)*
- Categorization Job
- Categorized Links

Visual Analytics

Analysis Server

Broadband Logs (IPDR)

- Unique Links to Crawl
- Unique Categorized Links
- Log data timestamp, IP...
- **Final Categorization Results** timestamp, IP, URL, categories

Data Store Categorization Results

Add customer profile, location, etc

§sas | THE POWER TO KNOW.

CHALLENGE | CATEGORIZING WEBSITES FROM LARGE IPDR DATA

- **Data size:**

  - 1 day of IPDR = 500 GB of data

  - 1 milliard URLs (10^9)

  - 1 million unique domains

- **Data to analyze:**

  - Only unique domains/ URLs are parsed and recorded, history from web proxy where possible

  - Top 100 domains make up 30% of all URLs – *taxonomy modeled from this sample*

  - >90% of URLs are advert/ monitoring sites – *sites excluded from parsing*

  - About 100m unique URLs to categorize per month – *# webpages to parse on 1st run, in 5 days*

    - *Mid size ECC server parses about 20m html docs per day (average 10kb text files, 4p4c cpu, 16gb memory)*

  - # webpages grow <10% per month (<10m pages) – *# webpages to parse on subsequent run, in <1 day*

 # Estimated from 50gb sample IPDR data

§SAS | THE POWER TO KNOW.

ENRICHED IPDR

INCLUDE WEBSITE CATEGORIZATION AND CUSTOMER PROFILE

Enriched IPDR

IPDR

- URL
- IP
- Date time

Categorization results

- Domain category
- Webpage category
- Sub-categories

Customer profile

- Location/ address
- Mobile/ household/ business
- Customer segment/ profile
- ...

§sas. | THE POWER TO KNOW.

SAMPLE RESULT | CATEGORIZATION RESULT DETAILS

All categories:

Top Category:

All categories that seem relevant to the URL – gives deeper insights to URL with multiple categories

Main Category:

The general category that is most relevant to the URL.

Top and sub categories, giving more detailed view

IP which generated the request

URL request

	ip	URL	Top Category	Main category	categories
250	116.14.159.71	http://www.asiawebdirect.com/	Economy	Economy, Business and Finance/T...	Top/Economy, Business and Fin...
251	116.14.83.61	http://itunes.apple.com/	Economy	Economy, Business and Finance/C...	Top/Economy, Business and Fin...
252	116.14.98.54	http://itunes.apple.com/	Economy	Economy, Business and Finance/C...	Top/Economy, Business and Fin...
253	116.14.147.54	http://www.groupon.sg/	Economy	Economy, Business and Finance/Fi...	Top/Economy, Business and Fin...
254	116.14.162.42	http://search.4shared.com/	Arts	Arts, Culture and Entertainment/La...	Top/Arts, Culture and Entertainm...
255	116.14.147.199	http://primeval.wikia.com/	Arts	Arts, Culture and Entertainment/Lit...	Top/Arts, Culture and Entertainm...
256	116.14.166.55	http://www3.nowgoal.com/	Sport	Sport	Top/Sport; Top/Sport/Basketball/...
257	219.74.243.233	http://www3.nowgoal.com/	Sport	Sport	Top/Sport; Top/Sport/Basketball/...
258	219.74.10.207	http://www.hbo.com/	Arts	Arts, Culture and Entertainment	Top/Arts, Culture and Entertainm...
259	116.14.34.110	http://sg.news.yahoo.com/	Sport	Sport	Top/Sport; Top/Economy, Busine...

Top/Arts, Culture and Entertainment; Top/Arts, Culture and Entertainment/Cinema; Top/Arts, Culture and Entertainment/Theatre/Music Theatre; Top/Economy, Business and Finance/Media/Satellite and Cable Service; Top/Sport/Boxing; Top/Arts, Culture and Entertainment/Entertainment (General); Top/Arts, Culture and Entertainment/Literature; Top/Arts, Culture and Entertainment/Arts (General); Top/Religion and Belief/Christianity /Anglican; Top/Crime, Law and Justice; Top/Social Issue/Family/Marriage

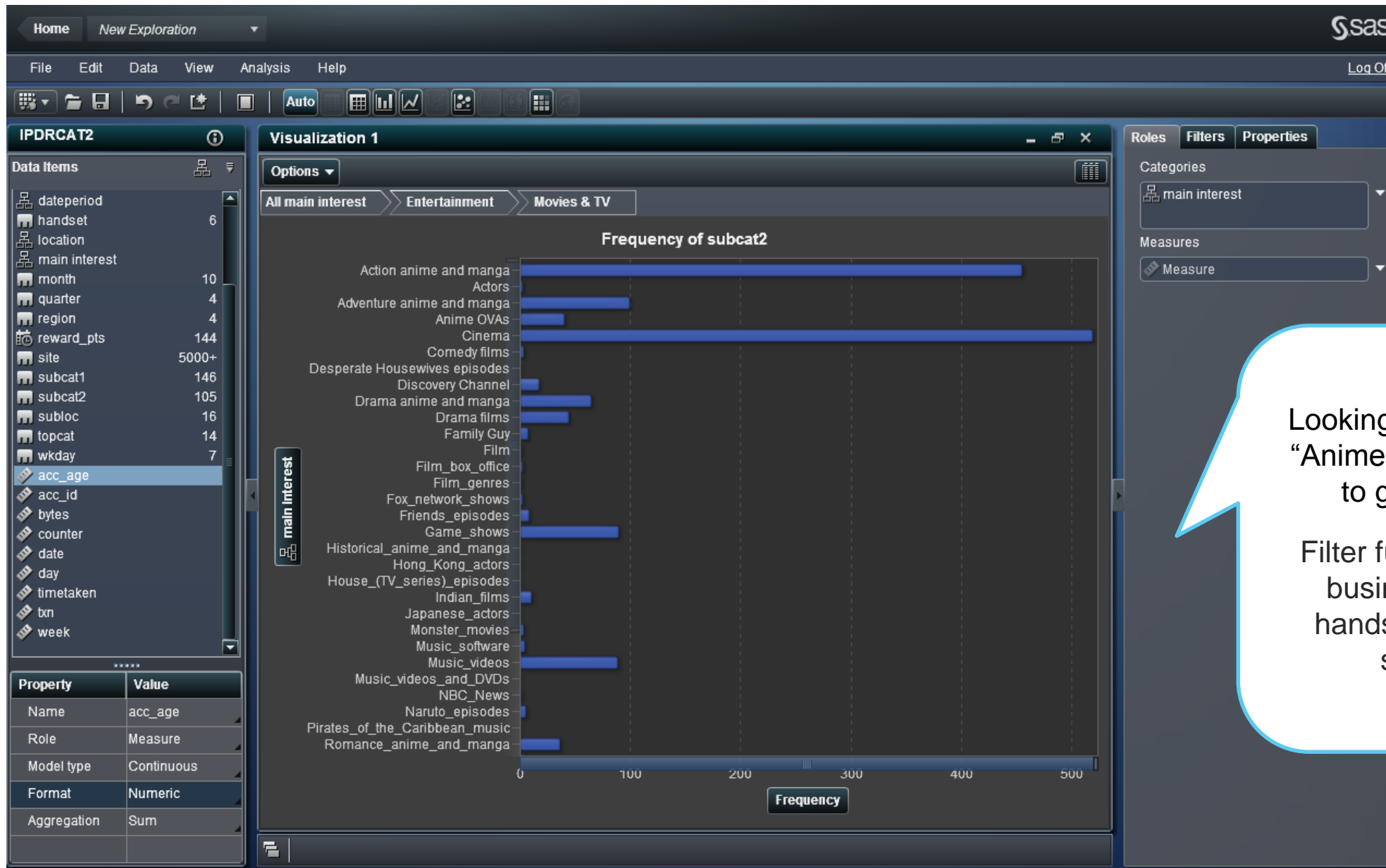| 263 | 219.74.50.21 | http://passport.game321.com/ | Arts | Arts, Culture and Entertainment/Int... | Top/Arts, Culture and Entertainm... |
| 264 | 116.14.236.226 | http://yugioh.wikia.com/ | Arts | Arts, Culture and Entertainment/Lib... | Top/Arts, Culture and Entertainm... |

§sas | THE POWER TO KNOW.

WHAT ARE PEOPLE MOST INTERESTED IN ?
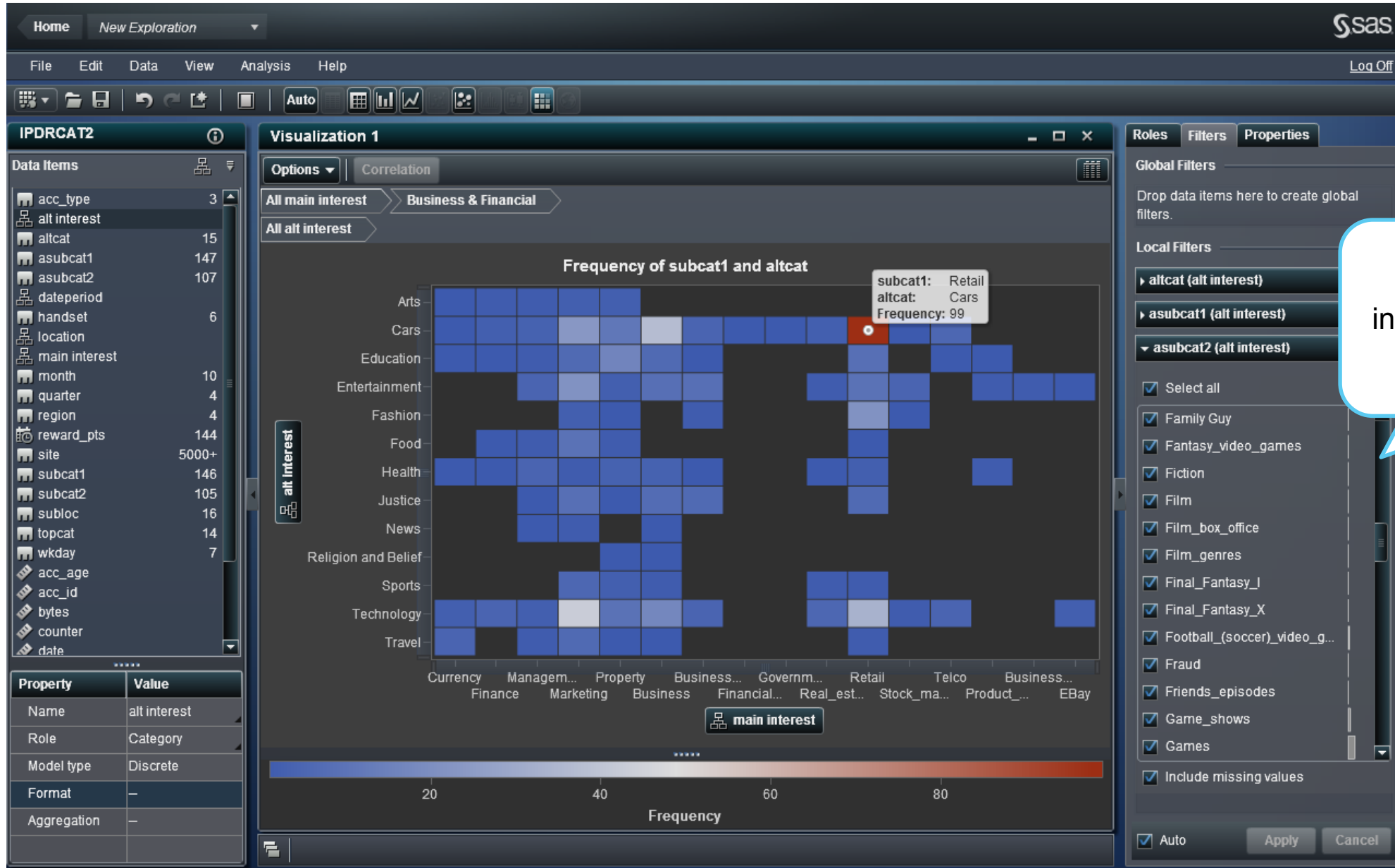## (DISTRIBUTION BY PAGE VIEWS)

WE CAN DEEP DIVE INTO AREAS OF INTEREST

Looking into Entertainment, "Anime" + "Cinema" seems to generate interest
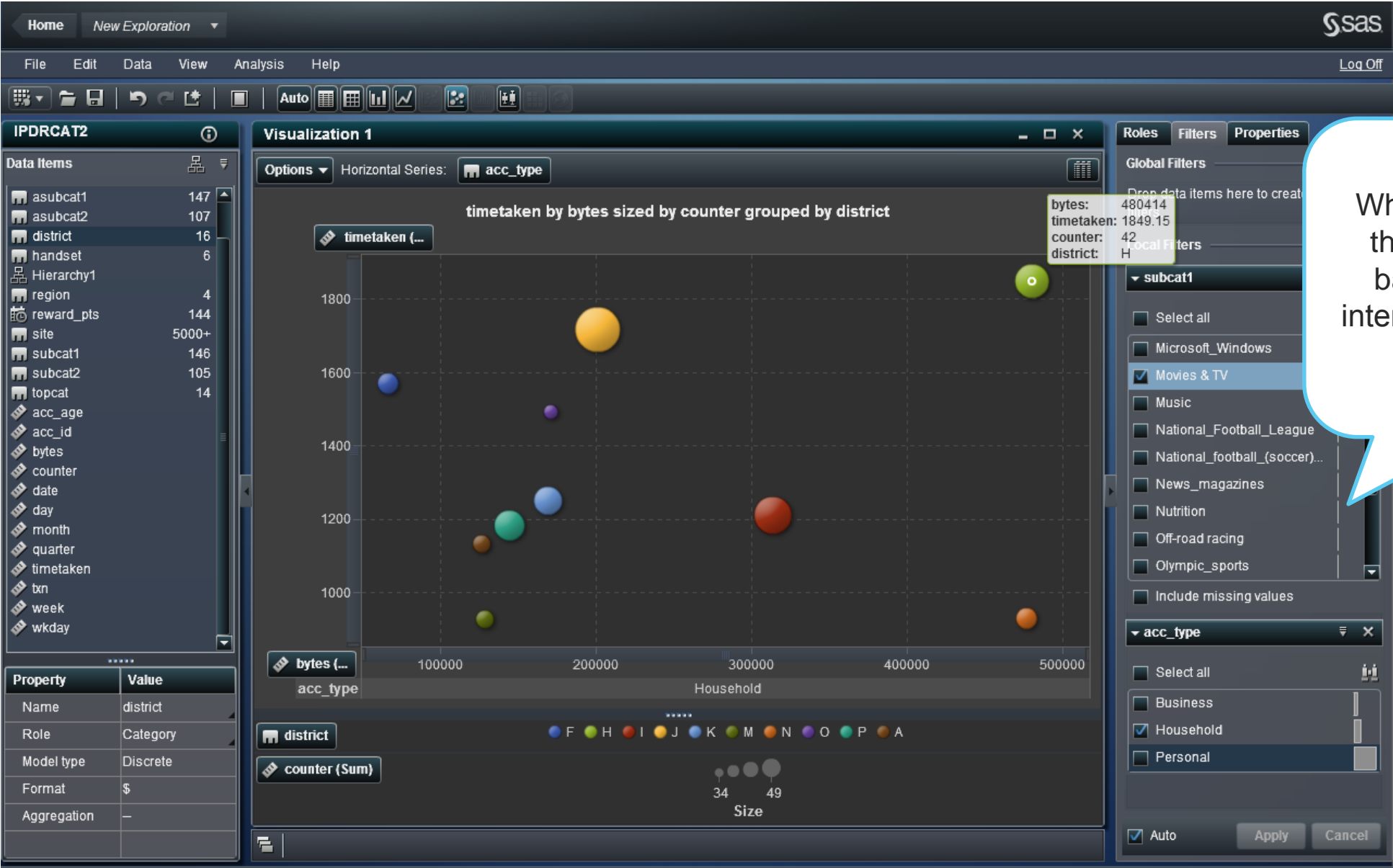
Filter further - by location/ business vs personal/ handset type/ customer segment/ etc.

WHAT ARE THE 2ND NATURE OF WEBSITES SURFED?

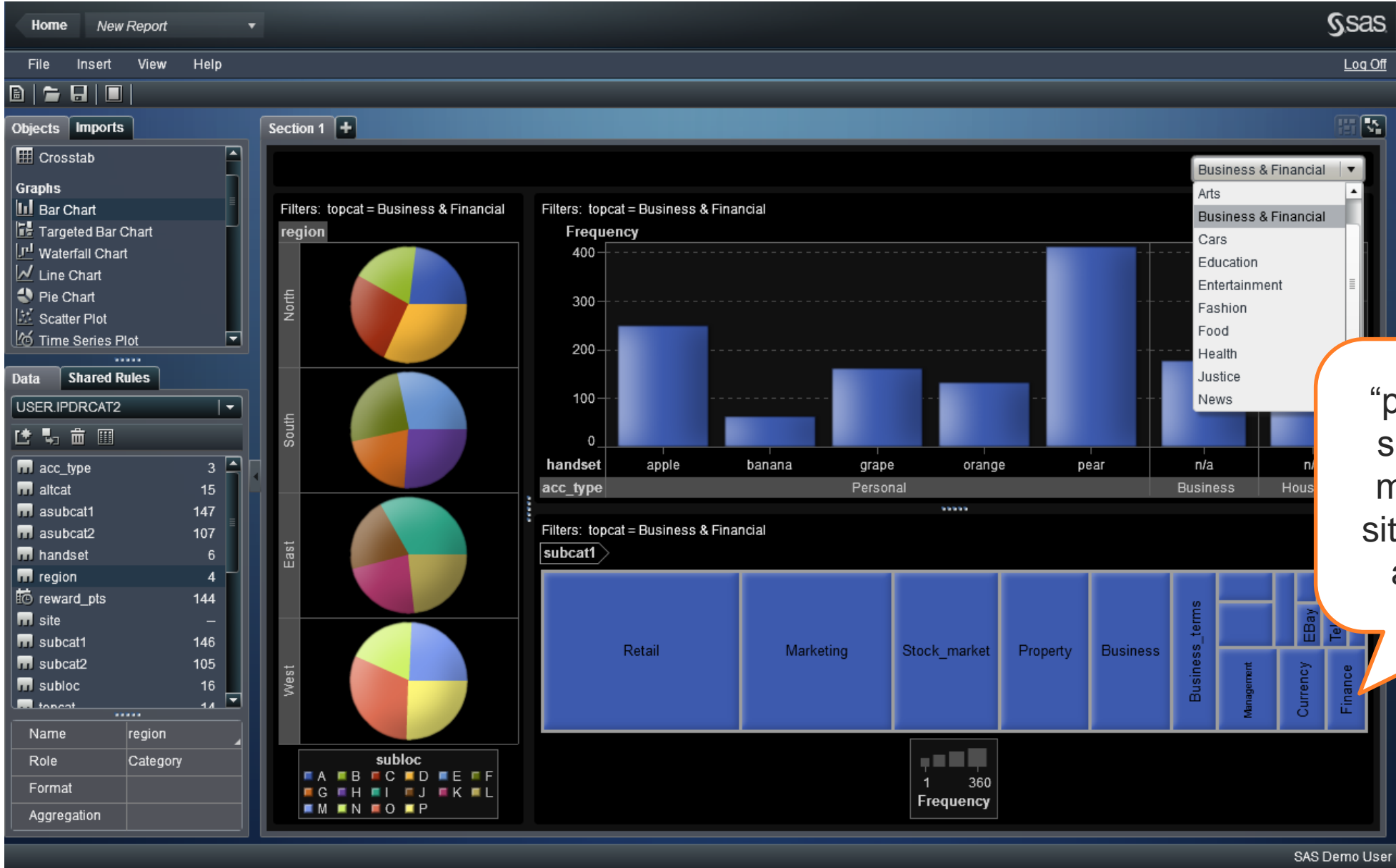Common/ strong interest in "Retail+Cars" related web pages

WHERE SHALL WE HOLD OUR "SPACE-AGE DRIVE-THRU MOVIE EVENT"?

While district "H" is not the largest customer base, it shows most interest in "Movies & TV" and "Sci-Fi"

WHAT IS THE PROFILE OF PEOPLE INTERESTED IN...
(BY REGION, HANDSET, SUB-INTERESTS, ETC)

"pear/ apple" users surf a lot of "retail/ marketing" related sites, evenly spread across locations

KEY BENEFITS | WHY CONSIDER ONLINE LIFESTYLE ANALYSIS?

1. Full, actual view of consumer interests, drillable into very specific areas:

   - What were households in neighborhood X interested in during evenings?

2. Immense IPDR is seldom analyzed for marketing/ customer behavior:

   CDR has limited/ no view of consumer lifestyles

   - Total data size downloaded vs % of consumers interested in coffee + hockey

3. SAS VA gives unparalleled speed/ flexibility to dive deep, far and wide:

   - No sampling, No reports to prepare, No surveys to run, No priority of big questions vs minor questions – all questions answered.

Customer Insights Module

ZapFi's intelligent mobile marketing platform which captures online surfing behaviour and creates detailed user profiles based upon SAS' tools and technologies.

Read more

ZAPFI: 'CUSTOMER INSIGHTS MODULE'

ZAPFI

BUSINESS GOALS

- ZapFi is a mobile marketing company based in Luxembourg that helps retail businesses with sophisticated personalized marketing through its Wi-Fi hotspot network
- Desire to create meaningful customer segments based on gradually built consumer profiles and in-store behavior
- Deliver real-time, relevant promotions and ads to consumers based on their segment membership and real-time behavior.

SOLUTION

- SAS® Data Management
- SAS® Visual Analytics
- SAS Text Analytics
- SAS® Enterprise Miner™
- SAS® Intelligent Advertising for Publishers
- SAS® Decision Manager

RESULTS

- Created detailed consumer segments based on Web browsing behavior and opt-in profiles.
- Using SAS® Visual Analytics provided customized insight reports and dashboards (via a web portal) to their customers and their stores focused on end-user consumer preferences, demographics and competitive intelligence.
- Deliver real-time, relevant ads and promotions while consumers engaged at ZapFi hotspots through SAS® Decision Manager and Intelligent Advertising functionality.
- Deliver personalized ads and promotions to consumers in real-time during the shopping experience
- Extend to include additional real-time data sources, like mobile telecommunications data

COMMUNICATIONS

"ZapFi enables sophisticated one-to-one mobile marketing so businesses can offer free Wi-Fi to their end-user customers. We also offer intelligent mobile marketing solutions that leverage the Wi-Fi channel. We needed a powerful real-time solution that could analyze all the data we collect from gradually built consumer profiles and online Web behavior."

Gery Pollet
Founder and Executive Chairman
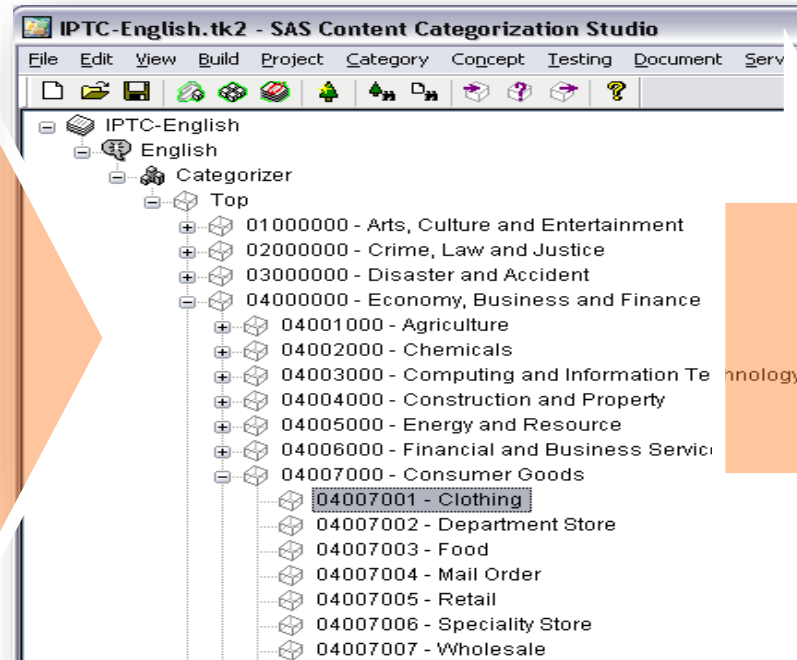ZapFi

Customer
Validated

§sas | THE POWER TO KNOW.

MAJOR TAIWAN CSP | ONLINE LIFESTYLE ANALYSIS

1. Explore consumer web surfing behavior in IP data records (IPDR)
2. Categorize web surfing behavior via a deep rooted taxonomy
3. Visualize the results over various dimensions to discover online lifestyles of consumers
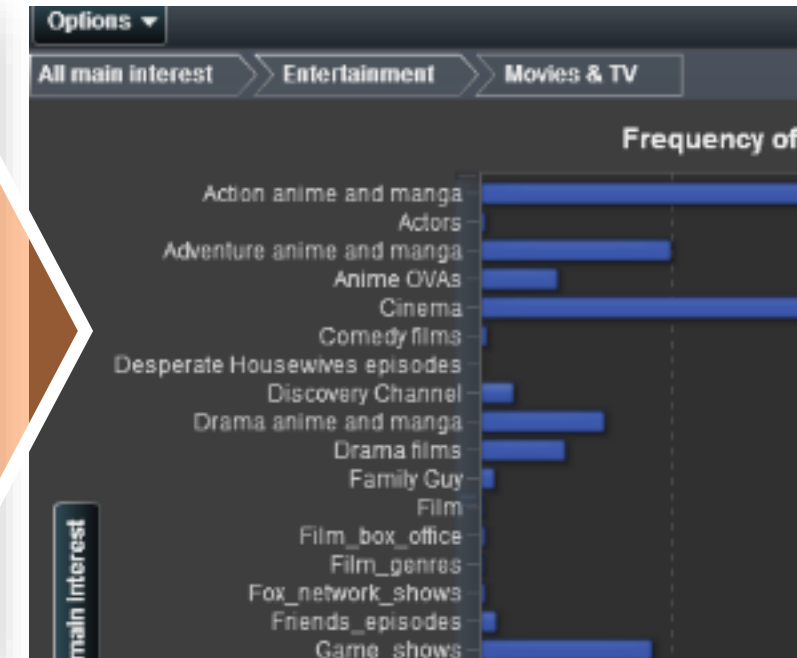
1. Collect URL content data

2. Categorize preferences

3. Visualize online lifestyle

§sas | THE POWER TO KNOW.

MAJOR TAIWAN CSP

MAPPING CUSTOMER'S 4G JOURNEY WITH SAS TEXT ANALYTICS ON SOCIAL MEDIA BIG DATA

Customer 4G adoption

Interaction

Awareness

- **Creating demand**
  - "Killer" apps/ content
  - CHT vs 3rd party

- **Learning about 4G**
  - Where? (news/ social/ sites)
  - How? (what is mentioned?)
  - Learning taxonomy
  - Top trends on 4G

- **Proxy (looking elsewhere)**
  - pattern of 4g adoption
  - 4g campaigns
  - Pos/ neg sentiments (SWOT)

Interest

- **Direct demand**
  - Search entries
    - Google/ yahoo/ etc
    - Search for products/ services/ technology
  - Discussion
    - Usage at home/ wifi
    - Content preference
- **Indirect demand**
  - Possible usage
    - Gaming/ video/ fileshare/ social media/ new apps
  - Complaints
    - Call center/ social media
    - Need for speed/
    - Sensitivity to speed issues

Adoption

- **Reasons for purchase**
  - Own needs
  - Social recommendation
  - Our marketing efforts

- **Competition**
  - Quality of service
  - Customer service
  - Contract/ offering
    - Price
    - Phone
    - Data
    - Policy

Evangelize

- **Profile of a champion**
  - Who loves us?
  - Key reasons?
  - Crowd response/ impact

- **Profile of an 'opposer'**
  - Who is strongly against us?
  - Key reasons?
  - Crowd response/ impact

WEB CATEGORIZATION - CONCEPT LEARNING

We collect sample data for each category – via wiki/ google/ udn…

Using SAS Text Analytics to analyze the samples for key concepts