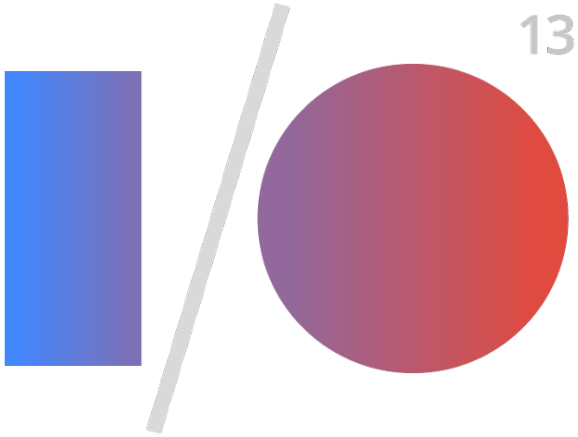
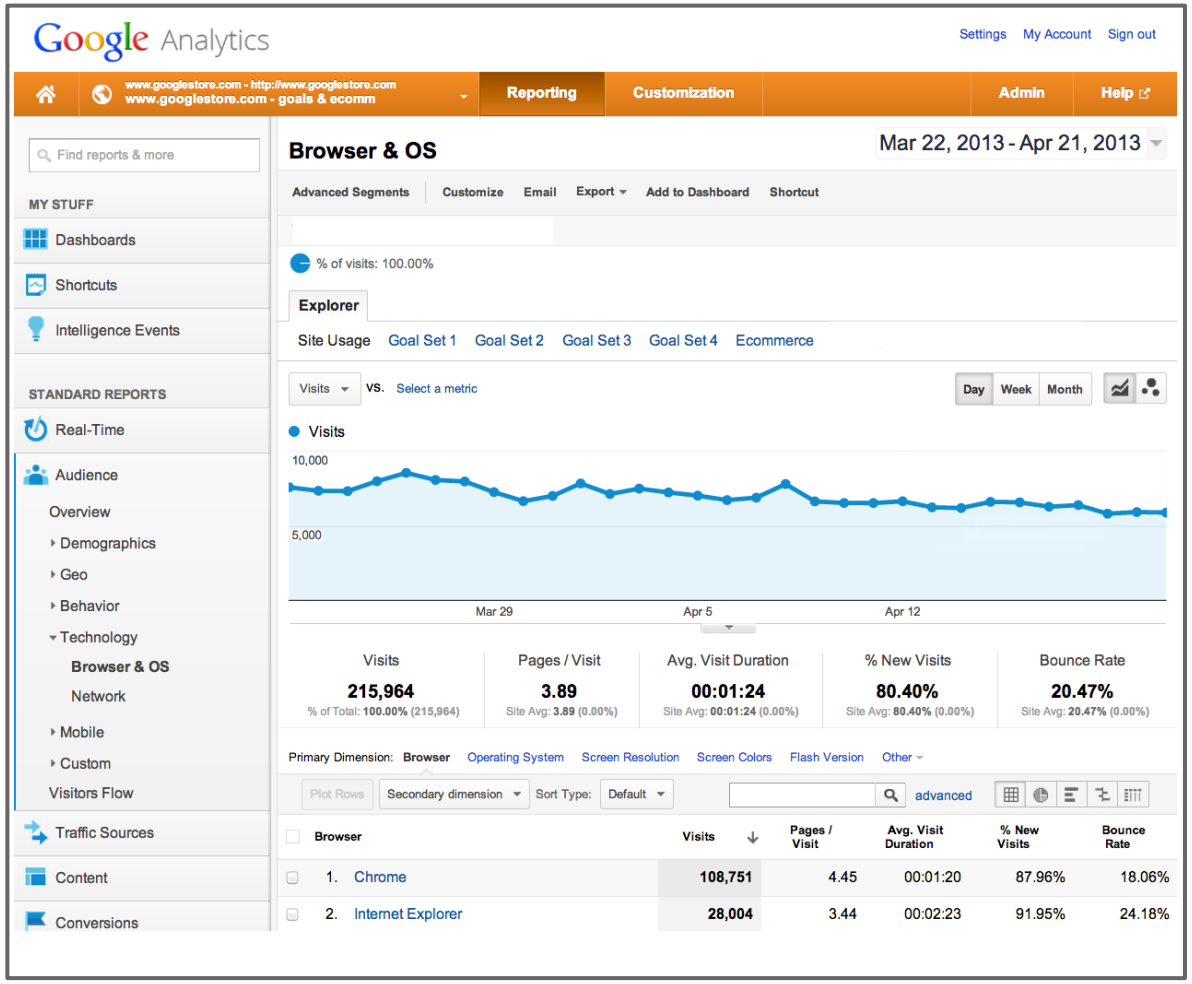

Google
Developers

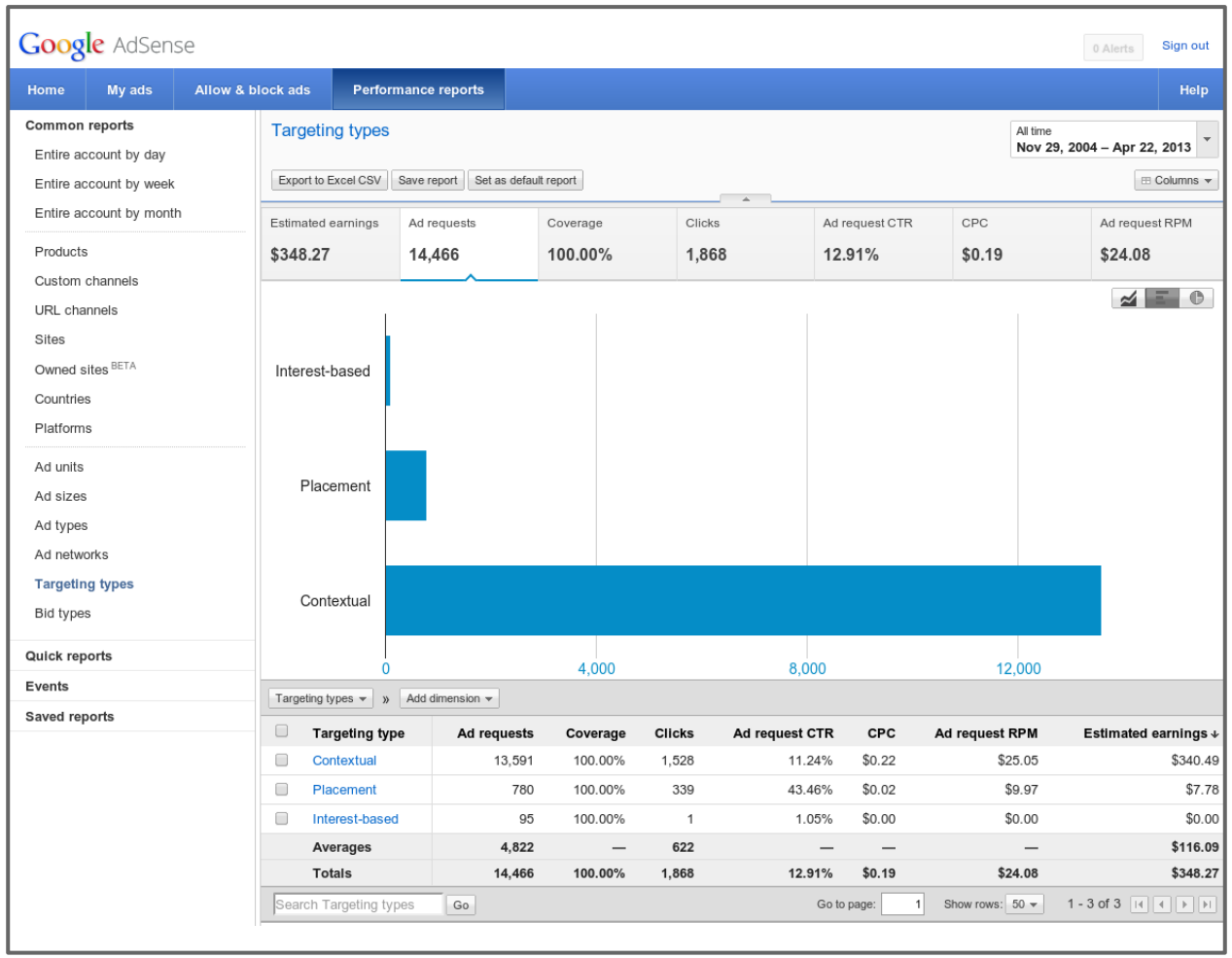
Google Analytics and AdSense Data Analysis in BigQuery

Louis Collard, Software Engineer, AdSense
Clancy Childs, Product Manager, Google Analytics



Designing Data Access for the "Average" User





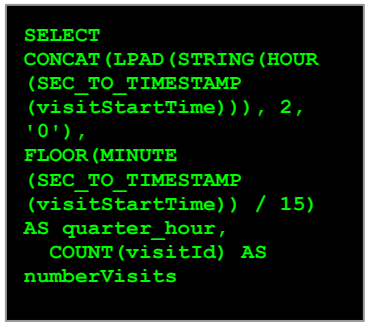
Designing Data Access for Data Scientists

Google Analytics

AdSense

+

Corporate Data (CRM, ERP)



Complex Ad-Hoc Queries

Reporting and Visualization

Data warehouses

Applications

BigQuery: Google's Big Data Analysis Engine

```
1  SELECT
2    date,
3    platform_type_name,
4    SUM(ad_requests),
5    SUM(earnings)
6  FROM
7    [google.com:adsense-reports:Reports.DailyReport]
8  WHERE
9    date > '2013-05-01'
10 GROUP BY
11   date,
12   platform_type_name;
```

RUN QUERY

Row	date	platform_type_name	ad_requests	earnings
1	2013-05-01	Tablets	1132581	3327.0
2	2013-05-01	High-end mobile devices	1904126	10241.0
3	2013-05-02	Desktop	7741152	24573.2
4	2013-05-02	Tablets	7474616	32643.8
5	2013-05-02	High-end mobile devices	5354473	13257.0

Huge Scale

Interactive Processing

Cost Effective

Third-Party Ecosystem

AdSense Data in BigQuery

AdSense Data in BigQuery

Tables

DailyReport

Daily**AdUnit**Report

Daily**CustomChannel**Report

Daily**UrlChannel**Report

Daily**Domain**Report

Common Dimensions

Date

Ad Client

Product

Ad Format

Bid Type

Targeting Type

Platform Type

Country

Metrics

Ad Requests

Matched Ad Requests

Individual Ad Impressions

Clicks

Earnings

Compared to the AdSense UI

Increased Flexibility
Request any combination of dimensions and metrics

Scalable
Query hundreds of thousands of rows

Live data
Matches the AdSense UI and API, updates throughout the day

Free to query
No charges for querying the AdSense BigQuery tables

Top Channels

BigQuery

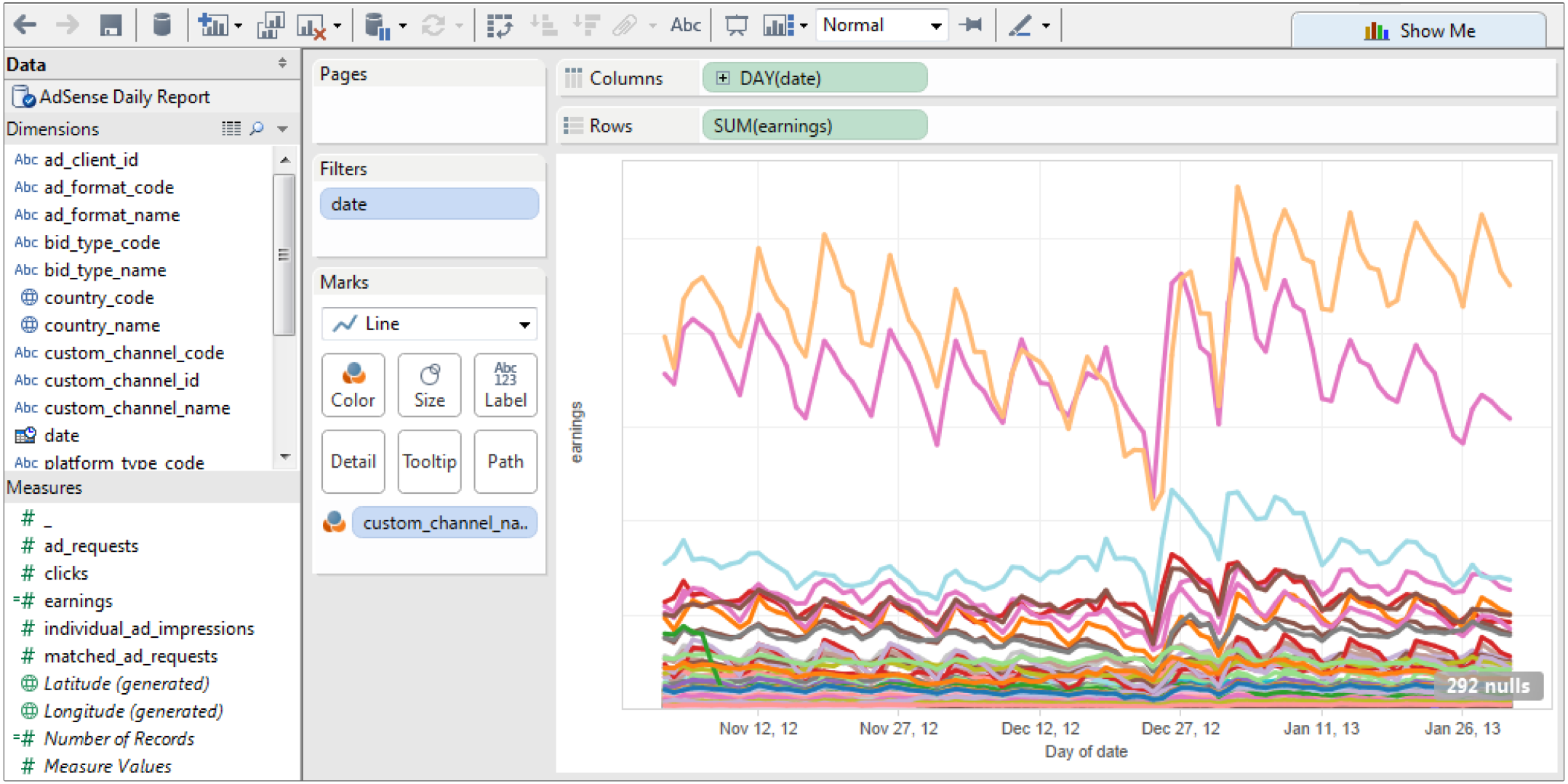
```
SELECT
 custom_channel_id,
 SUM(earnings) AS earnings
FROM
 [google.com:adsense-reports:Reports.DailyCustomChannelReport]
GROUP BY
 custom_channel_id
ORDER BY
 earnings DESC
LIMIT
 5
```

Top Growing Channels

BigQuery

```
SELECT
  jan_report.custom_channel_name,
  feb_report.clicks_sum - jan_report.clicks_sum AS clicks_delta
FROM
  (SELECT custom_channel_name, SUM(clicks) AS clicks_sum
   FROM [...DailyCustomChannelReport]
   WHERE date  >= '20130101' AND date < '20130201'
   GROUP BY custom_channel_name) AS jan_report
INNER JOIN
  (..WHERE date  >= '20130201' AND date < '20130301'..) AS feb_report
ON jan_report.custom_channel_name = feb_report.custom_channel_name
ORDER BY clicks_delta DESC LIMIT 5
```

Third party integration: Tableau



Getting Started

Beta program open to all AdSense publishers from today!

Set up a BigQuery project for your AdSense user

http://bigquery.cloud.google.com

Start querying data

http://developers.google.com/adsense/bigquery

For support or to report problems

http://developers.google.com/adsense/community

Google Analytics and BigQuery

Google Analytics Data in BigQuery

What are we announcing today?

- Google Analytics data will be made available in BigQuery for Google Analytics Premium clients.
- Targeting a full launch by September 2013.
- Opening up a group for developers that are interested in working with this data when it is available.

Google Analytics Data in BigQuery

- A daily export from a Google Analytics profile into a BigQuery dataset.
- Each row represents a session/visit in the Google Analytics profile.
  - Session-level data - traffic source, custom dimensions, etc.
  - Hit-level data - sequenced hits included within it's session/row
- Dimensions and metrics map to the names used in the Google Analytics Core Reporting API.
- At launch, this will contain first-party (site) data.
  - Some post-processed data like Geography or AdWords may not be available at launch.

Google Analytics Data in BigQuery - Developer's Group

For information about
Google Analytics Premium:

http://google.com/analytics/premium

For future updates, please complete the form:

http://goo.gl/QJR9Y

Duncan McKie

Senior Business Analyst
eBay Classifieds Group

eBay Classifieds Group - Our Business

- Innovative leader in online classifieds, with 10 brands worldwide.
- 42m monthly users
- Connect buyers and sellers in more than 1,000 cities around the globe.

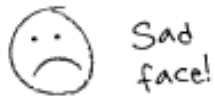
eBay Classifieds Group - Our Data

- Petabytes of internal data:
  - site databases
  - CRM
  - data warehouses
  - marketing systems
- External data:
  - Google Analytics
  - AdSense
  - payment providers
  - advertising data (DoubleClick, AdWords, other networks)

eBay Classifieds Group - Our Challenges

- Integrating data from different providers and making it available to large numbers of analysts across the business.
- Evaluating large volumes of data to investigate complex issues.
- Providing an accessible source of insight to enable non-power users to make data-driven decisions.

404 Page not found

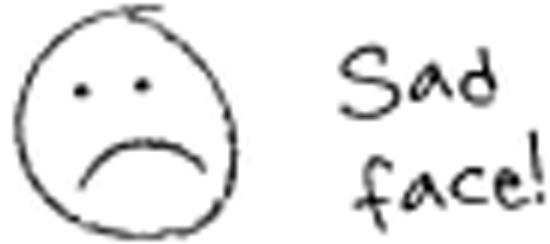


Sorry, the page you're trying to load doesn't seem to exist.

Back to Gumtree.com

Debugging Errors

- Errors such as 404 & 500 lose users!
- Difficult to find the cause of errors using Google Analytics UI
- Hit-level data in BigQuery allows us to examine user journeys



Debugging Errors: Step 1

BigQuery

```
SELECT
  visitorId AS visitorId1
  ,visitId AS visitId1
  ,hits.customVariables.customVarValue AS errorType
  ,hits.hitNumber
FROM
  [gumtree-uk-ga:google_analytics.sessions_20130421] AS t1
WHERE
  hits.customVariables.index = 1
  AND REGEXP_MATCH(hits.customVariables.CustomVarValue,'404|500');
```

Debugging Errors: Step 2

BigQuery

```
SELECT
  errorType.visitorId1
  ,errorType.visitId1
  ,errorType.errorType
  ,MAX(previousHit.hits.hitNumber) AS hitNumber
FROM
  (SELECT
    ...
    ) AS errorType
INNER JOIN EACH
    (FLATTEN ([gumtree-uk-ga:google_analytics.sessions_20130421], hits.hitNumber))
        AS previousHit
    ON errorType.visitorId1 = previousHit.visitorId
    AND errorType.visitId1 = previousHit.visitId
WHERE
  errorType.hits.hitNumber < previousHit.hits.hitNumber
GROUP EACH BY 1,2,3;
```

Debugging Errors: Step 3

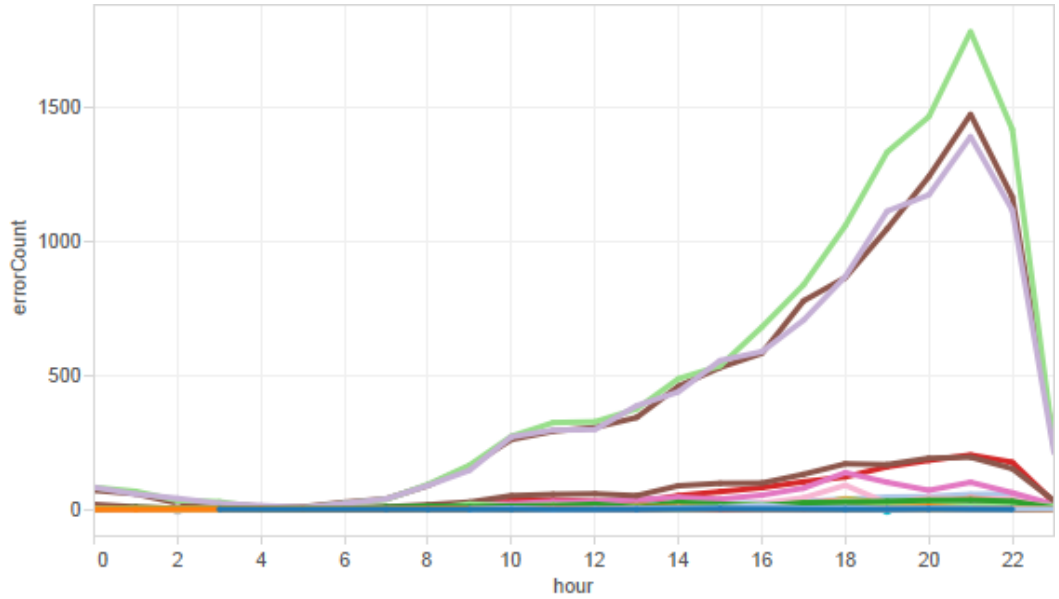
BigQuery

```
SELECT
  pageName.hits.hour AS hour
  ,hitFinder.errorType AS errorType
  ,pageName.hits.customVariables.customVarValue AS pageName
  ,COUNT(*) AS errorCount
FROM
     (FLATTEN([gumtree-uk-ga:google_analytics.sessions_20130421],hits.hitNumber)) AS pageName
INNER JOIN EACH
     (SELECT
          ...
     ) AS hitFinder
     ON pageName.visitorId = hitFinder.visitorId1
     AND pageName.visitId = hitFinder.visitId1
     AND pageName.hits.hitNumber = hitFinder.hitNumber
WHERE
  pageName.hits.customVariables.index = 1
GROUP EACH BY 1,2,3
ORDER BY 4 DESC;
```

Debugging Errors: The Results

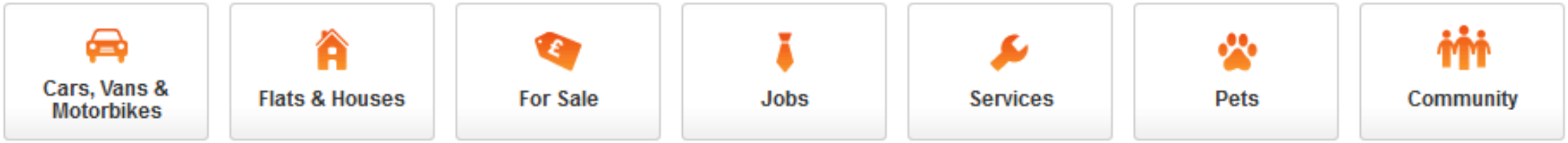
Query Results

Row	hour	errorType	pageName	errorCount
1	21	Error-404	Search	1775
2	21	Error-404	VIP	1469
3	20	Error-404	Search	1458
4	22	Error-404	Search	1416
5	21	Error-404	Listing	1383



- Can investigate any sequence of hits
- No need to set up funnels in advance

Segmenting Onsite Activity



- Each post has attributes (category, location, etc.) used across business to segment user activity

- Segments can be complex and dynamic, changing with business needs

Segmenting by TV Region: Step 1

- Import lookup tables into BigQuery as CSV or JSON

Row	location_id	location_level	location_level_1	location_level_2	location_level_3	location_level_4	location_name
1	881	4	Midlands	West Midlands	Birmingham	Hodge Hill	Hodge Hill
2	870	4	Midlands	West Midlands	Birmingham	Tile Cross	Tile Cross
3	546	3	South	East Sussex	Brighton	null	Brighton
4	905	4	Midlands	West Midlands	Birmingham	West Bromwich	West Bromwich
5	836	4	Midlands	Nottinghamshire	Nottingham	Dunkirk	Dunkirk
6	939	4	Midlands	West Midlands	Birmingham	Winson Green	Winson Green

Segmenting by TV Region: Step 2

BigQuery

```
SELECT
   hits.customVariables.customVarValue AS location_level_2
   ,visitId
   ,visitorId
   ,COUNT(hits.hitNumber) AS pageviews
   ,SUM(totals.bounces) AS bounces
FROM
   [gumtree-uk-ga:google_analytics.sessions]
WHERE
   hits.customVariables.index = 4
GROUP EACH BY 1,2,3;
```

Segmenting by TV Region: Step 3

BigQuery

```
SELECT
   location_level_2
   ,location_level_1 AS tv_region
FROM
   [gumtree-uk-ga:gumtree_uk_metadata.location]
WHERE
   location_level_1 IN ('North West',
   'South',
   'Midlands',
   'South West')
GROUP BY 1,2
```

Segmenting by TV Region: Step 4

BigQuery

```
SELECT
    location.tv_region
    ,COUNT(DISTINCT CONCAT(STRING(ga.visitorId),'-',STRING(ga.visitId))) AS visits
    ,COUNT(DISTINCT ga.visitorId) AS visitors
    ,SUM(ga.pageviews) AS pageviews
    ,SUM(ga.bounces) AS bounces
FROM
    (SELECT
        ...
     ) AS ga
INNER JOIN
    (SELECT
        ...
     ) AS location
    ON ga.location_level_2 = location.location_level_2
GROUP BY 1;
```

Segmenting by TV Region: The Results

Query Results

Row	location_location_level_1	visits	visitors	pageviews	bounces
1	North West	137010	99368	672986	25279
2	South	110954	85544	523778	20578
3	Midlands	217683	161621	1105173	38502
4	South West	93297	70974	445084	16032

- No need to individually query each segment
- No pre-aggregation or sampling of data

A|B Testing - AdSense & Analytics

- A|B testing allows us to improve our implementation
- Interested in impact on whole user experience
- Optimisation metrics (e.g. Clicks/Visit) derived from various sources

A|B Testing - AdSense & Analytics: Step 1

BigQuery

```
SELECT
   custom_channel_name AS adSenseGroup
   ,SUM(ad_requests) AS ad_requests
   ,SUM(matched_ad_requests) AS matched_ad_requests
   ,SUM(clicks) AS clicks
   ,SUM(earnings) AS earnings
FROM
   [874385582184:Reports.DailyCustomChannelReport]
WHERE
   REGEXP_MATCH(ad_client_id,'gumtree-com')
GROUP BY 1
```

A|B Testing - AdSense & Analytics: Step 2

BigQuery

```
SELECT
   hits.customVariables.customVarValue AS adSenseGroup
   ,COUNT(DISTINCT CONCAT(STRING(visitorId),'-',STRING(visitId))) AS visits
   ,COUNT(DISTINCT visitorId) AS visitors
   ,COUNT(hits.hitNumber) AS pageviews
FROM
   [gumtree-uk-ga:google_analytics.sessions]
WHERE
   hits.customVariables.index = 42
GROUP BY 1
```

A|B Testing - AdSense & Analytics: Step 3

BigQuery

```
SELECT
    ga.adSenseGroup
    ,ga.visits
    ,adSense.matched_ad_requests
    ,FLOAT(adSense.clicks/adSense.matched_ad_requests) AS CTR
    ,FLOAT(adSense.matched_ad_requests/ga.visitors) AS matched_ad_requests_per_visitor
    ,FLOAT(ga.pageviews/ga.visits) AS pageviews_per_visit
FROM
    (SELECT
    ...
    ) AS ga
INNER JOIN
    (SELECT
    ...
    ) AS adSense
    ON ga.adSenseGroup = adSense.adSenseGroup
```

A|B Testing - AdSense & Analytics: The Results

Query Results

Row	test_group	CTR	matched_ad_requests_per_visitor	pageviews_per_visit
1	testGroup1	0.020860557309110735	7.107795596644189	9.720080938168486
2	testGroup2	0.022982172537055166	7.505981406371053	10.323799771939214
3	testGroup3	0.023248034788426 16	7.078607463976023	10.050262812089356

- Able to evaluate all factors in one place
- Using same Custom Dimension means one setup for all tests

In Summary / Next Steps

Get up and running with BigQuery: **http://bigquery.cloud.google.com**

Google Analytics in BigQuery

- Register for more information: **http://goo.gl/QJR9Y**
- Targeting September Launch for Google Analytics Premium users

AdSense in BigQuery

- Beta available now!
- Step-by-step instructions on getting started: http://developers.google.com/adsense/bigquery

Thank You and Questions?

BigQuery: **http://bigquery.cloud.google.com**

AdSense: **http://developers.google.com/adsense/bigquery**

Google Analytics Premium Data: **http://goo.gl/QJR9Y**


Google
Developers