

Understanding Documentation Use Through Log Analysis

An Exploratory Case Study of Four Cloud Services



Daye Nam

dayen@cs.cmu.edu



Andrew Macvean

amacvean@google.com



Brad Myers

bam@cs.cmu.edu



Bogdan Vasilescu

vasilescu@cmu.edu



Carnegie Mellon University
School of Computer Science



Learning to use software services is hard

Lots of information is scattered across multiple pages



?

Cloud SQL Admin API

API for Cloud SQL instances
Enterprise Plus Edition

Service: sqladm Johannesburg (africa-south1)

Service endpoint

A [service endpoint](#) (<https://sqladm:8080>) specifies the network address for this service. This service has the following endpoint:

- <https://sqladm:8080>

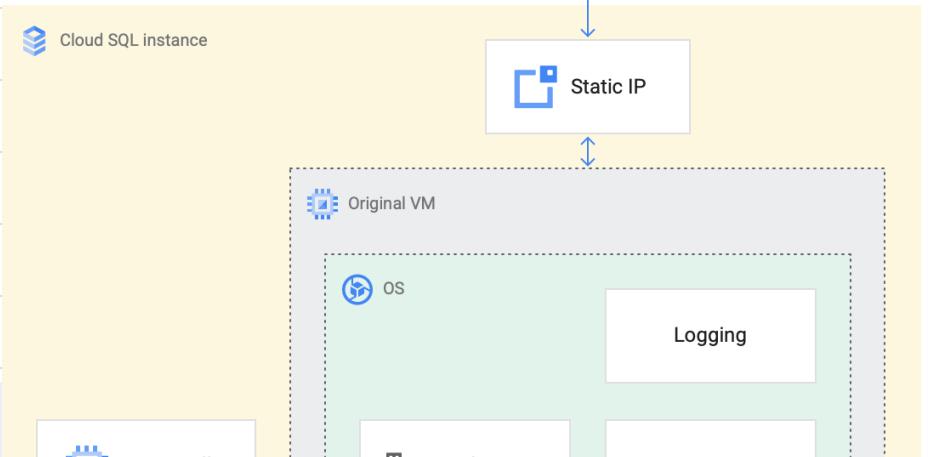
vCPUs
Memory
HA vCPUs
HA Memory
Data Cache Storage
HA Data Cache Storage

REST Reference

Pricing

Hourly Monthly

Product Overview



Cloud SQL instance

Original VM

OS

Logging

Database

Monitoring

HA Standard

Learning to use software services is hard

More challenging if a user is new to the service or application domain



Experienced
SQL Engineer



REST Reference

Cloud SQL Admin API

API for Cloud SQL instance management.

Service: sqladmin.googleapis.com

Service endpoint

A [service endpoint](#) (https://cloud.google.com/apis/design/glossary#api_service_endpoint) specifies the network address of an API service. One service might have multiple endpoints. This service has the following service endpoint and all URLs below are relative to it.

- <https://sqladmin.googleapis.com>

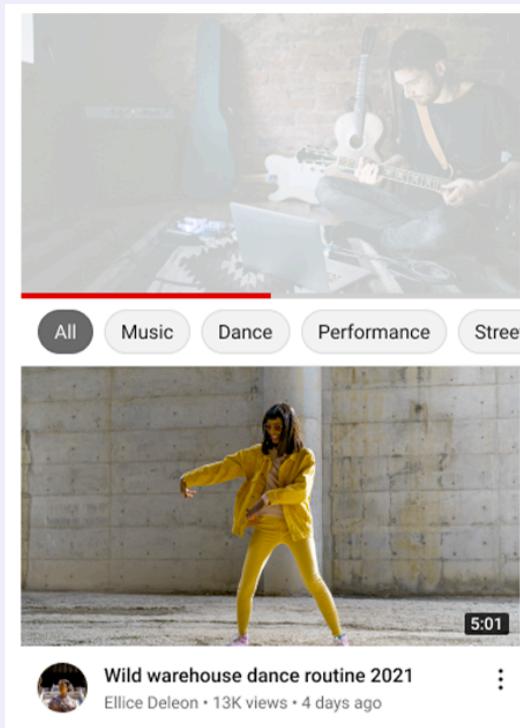
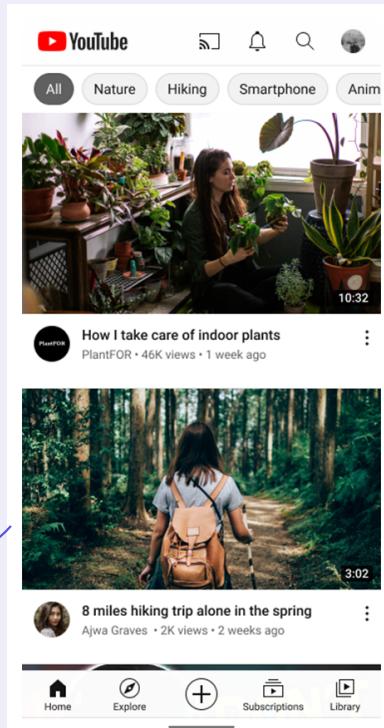


New Grad SWE
New to SQL



What if, we can recommend documentation like how Spotify or YouTube does?

*Youtube
recommendations
on homepage.*



*Recommendations
on "Up next."*

<https://blog.youtube/inside-youtube/on-youtubes-recommendation-system/>

What if, we can recommend documentation like how Spotify or YouTube does?



REST Reference

Cloud SQL Admin API

API for Cloud SQL instance management.

Service: sqladmin.googleapis.com

Service endpoint

A [service endpoint](#) (https://cloud.google.com/apis/design/glossary#api_service_endpoint) specifies the network address of an API service. One service might have multiple endpoints. This service has the following service endpoint and all URLs below are relative to it.

- <https://sqladmin.googleapis.com>

What if, we can recommend documentation like how Spotify or YouTube does?

SQL Key Terms:

To use Cloud SQL effectively, you should understand ...

Cloud SQL instance

A Cloud SQL instance corresponds to one virtual machine (VM). The VM includes the database instance and accompanying software containers to keep the database instance up and running.

Database instance

A database instance is the set of software and files that operate the databases: MySQL, PostgreSQL or SQL Server.

Public and private IP

In Cloud SQL, public IP means that the instance is accessible through the public internet. In contrast,

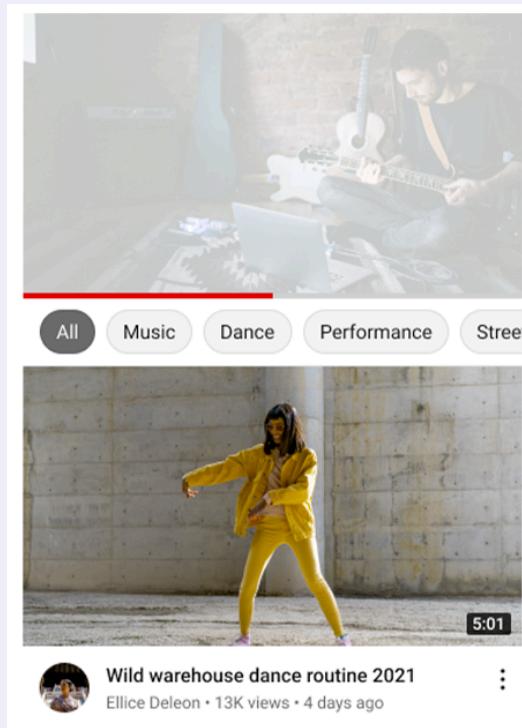
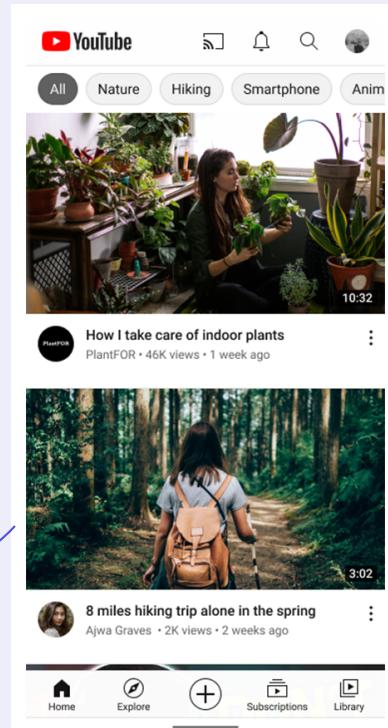
!

New-grad SWE
New to SQL



How can we infer users' info. needs for documentation recommendations?

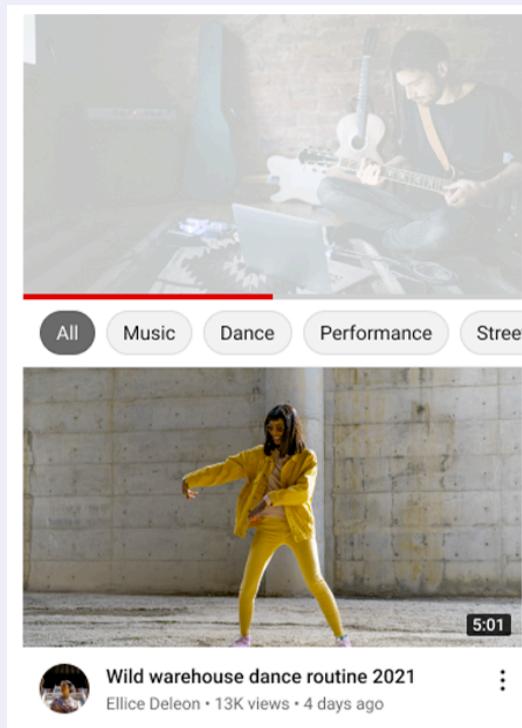
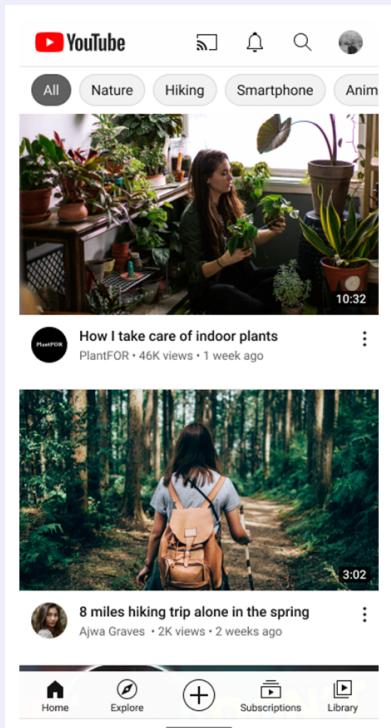
*Youtube
recommendations
on homepage.*



*Recommendations
on "Up next."*

<https://blog.youtube/inside-youtube/on-youtubes-recommendation-system/>

Many recommendation systems use signals from interaction logs



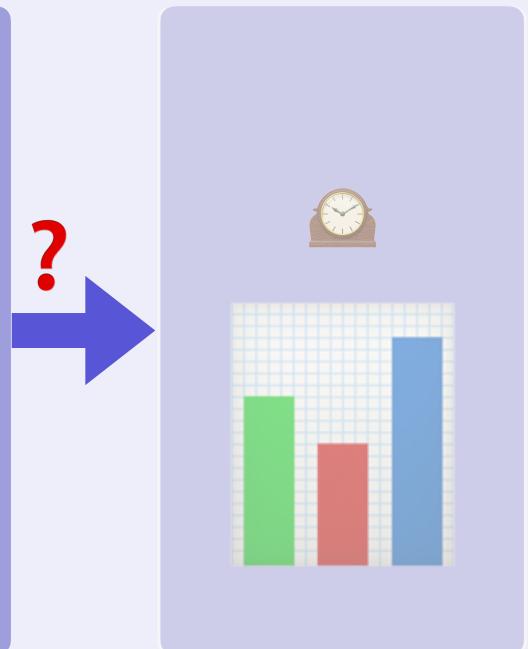
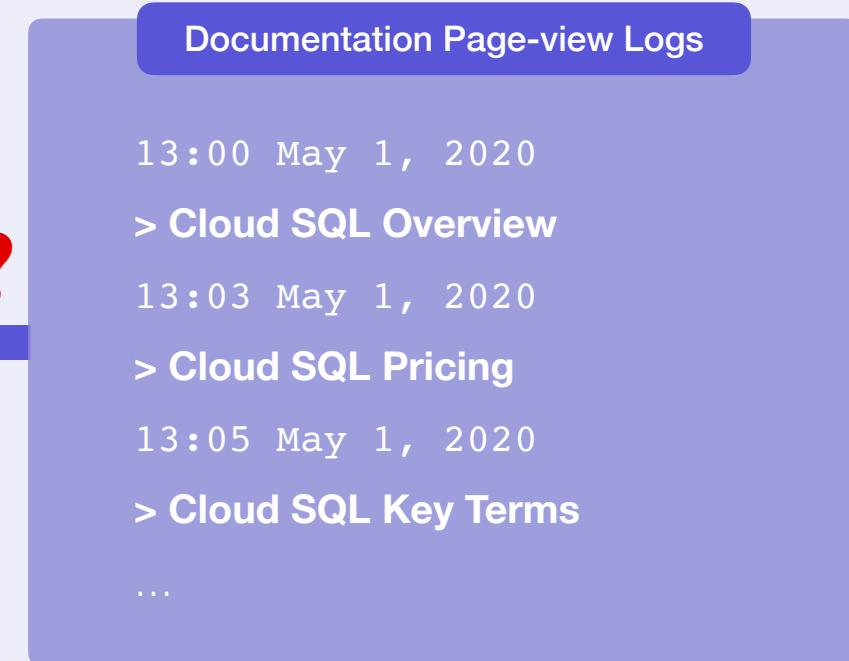
On YouTube's recommendation system

A **number of signals** build on each other to help inform our system about what you find satisfying: **clicks, watchtime, survey responses, sharing, likes, and dislikes.**

<https://blog.youtube/inside-youtube/on-youtubes-recommendation-system/>

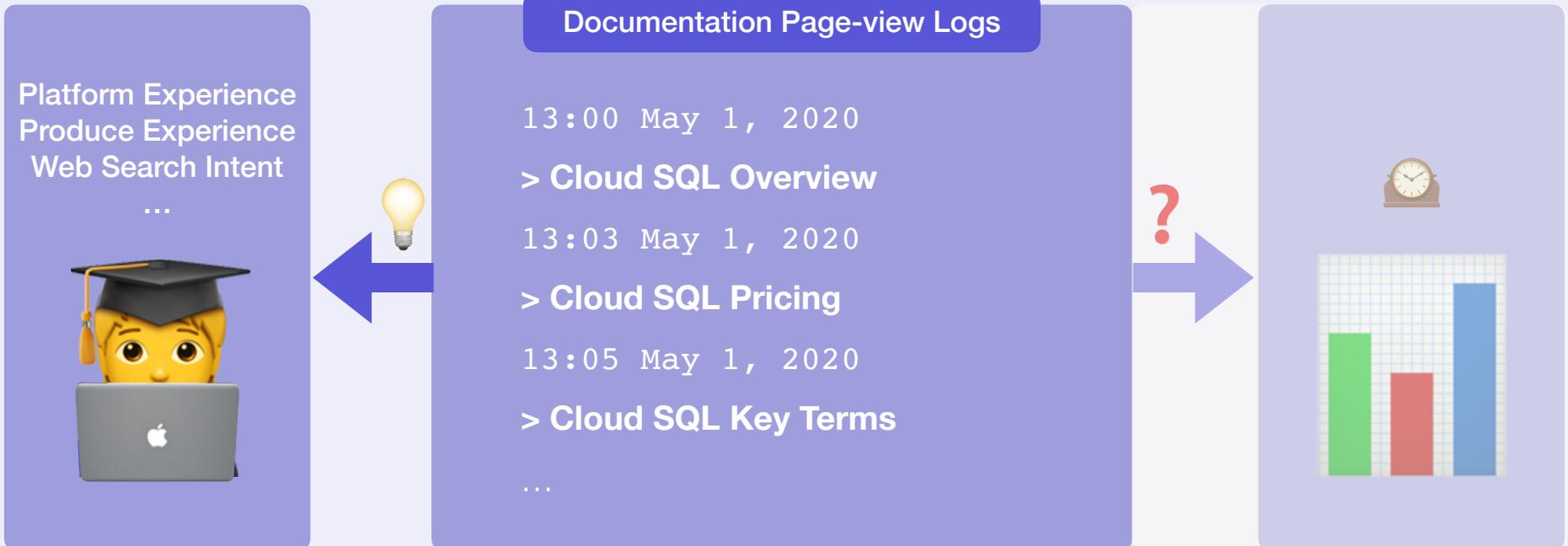
***Can documentation page-view logs be used
to estimate user characteristics or intent?***

Can documentation page-view logs be used to estimate user characteristics or intent?



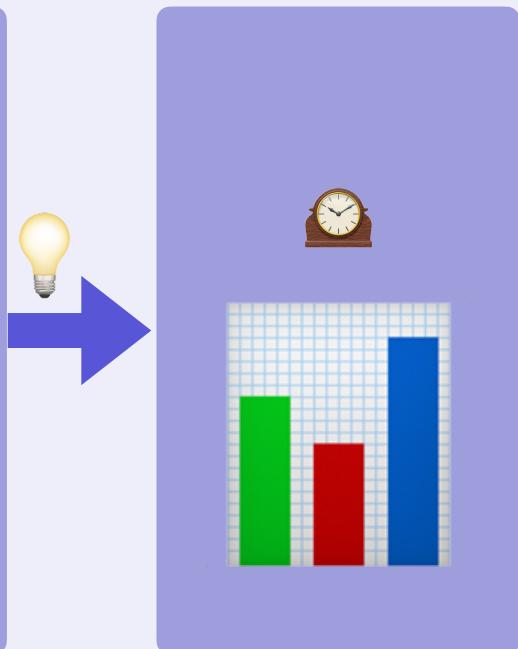
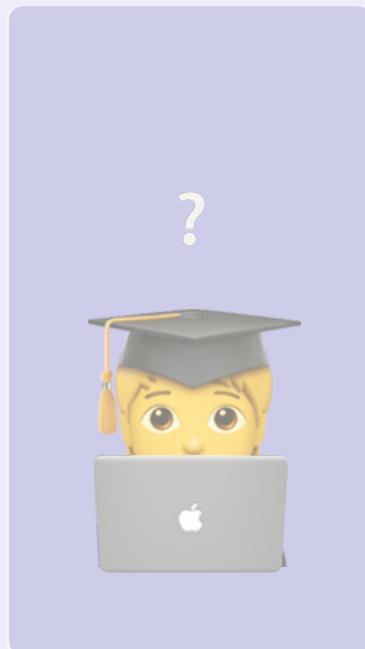


Documentation usage in page-view logs can be used in estimating user characteristics





Documentation usage in page-view logs can be used in estimating future product adoption





Google Cloud page-view logs analysis



Google Cloud page-view logs analysis



13:00 May 1, 2020
> Cloud SQL Overview
13:03 May 1, 2020
> Cloud SQL Pricing



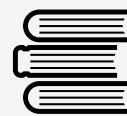
of more than **100,000 users**
of 4 Google Cloud products,
of May 2020 (aggregated)
with **privacy-preserving** techniques

Documentation Page-view Logs

User	Product	Dwell Time (m)				Pricing
		Reference	How-to	
0	P1	1	0	0
1	P1	35	0	0
:	:	:	:	:	:	:

Can documentation page-view logs be used to estimate user characteristics or intent?

Hypotheses Building



Literature



H1-5

H1: Experience

Earl 15

Maalej 14

H2: Product Type

Freund 15

Montesi 08

H3: Doc. Type Predisposition

Meng 18

H4: Possible Intent

Brandt 09

Rao 20

H5: Subsequent API Use

Vargas 20

Rauf 16

High experience levels are positively associated with accessing documentation covering implementation details

Using page-view logs & API usage data for hypothesis testing

Page-view Logs

API Usage data

H1: Experience

← Account Age (years), Past Succ. Req

Using page-view logs & API usage data for hypothesis testing

Page-view Logs

API Usage data

H3: Doc. Type Predisposition ←Page views in the previous three months

Using page-view logs & API usage data for hypothesis testing

Page-view Logs

API Usage data

H1: Experience

← Account Age (years), Past Succ. Req

H2: Product Type

← Product

H3: Doc. Type Predisposition

← Page views in the previous three months

H4: Possible Intent

← Average per-page dwell time

H5: Subsequent API Use

← Number of successful API requests

Using page-view logs & API usage data for hypothesis testing

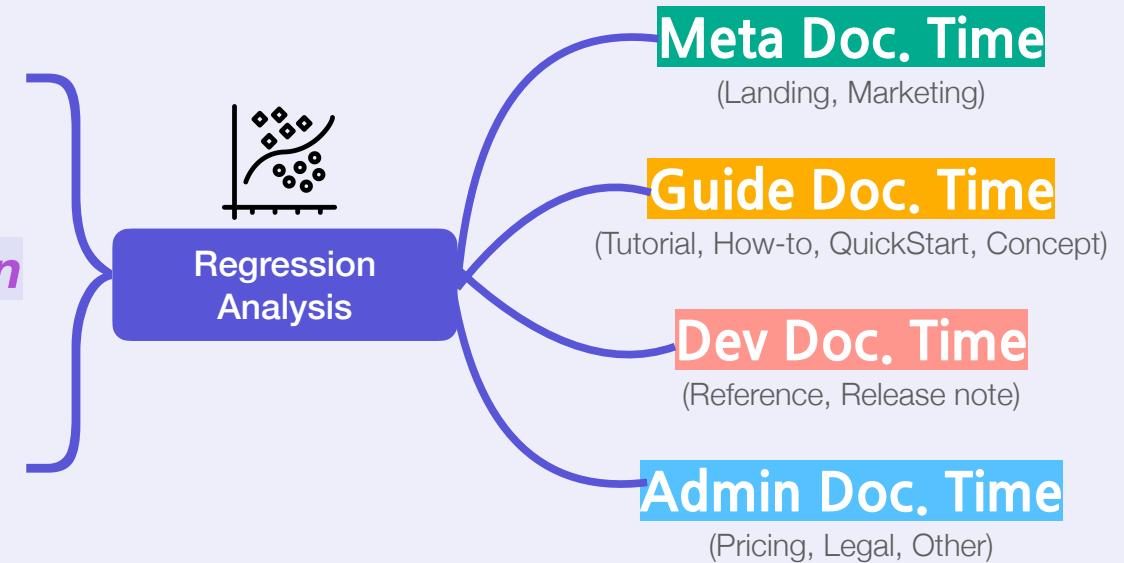
H1: Experience

H2: Product Type

H3: Doc. Type Predisposition

H4: Possible Intent

H5: Subsequent API Use



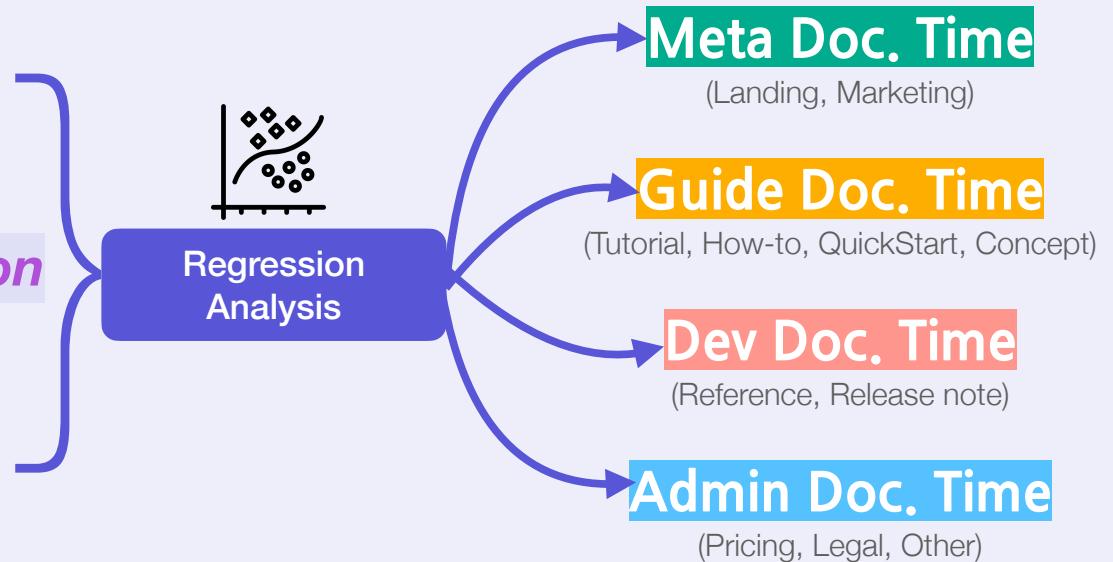


*Documentation usage in page-view logs
correlates with user characteristics &
future product adoption*



Documentation usage in page-view logs correlates with user characteristics & future product adoption

- ✓ **H1: Experience**
- ✓ **H2: Product Type**
- ✓ **H3: Doc. Type Predisposition**
- !? **H4: Possible Intent**
- ✓ **H5: Subsequent API Use**



High experience levels are positively associated with accessing Dev documentation

H1: Experience

Product Experience
(Successful API requests
in prev. 3 months > 0)



Regression
Analysis

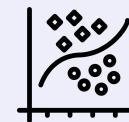
+ (1.34)

Dev Doc. Time > 0
(Reference, Release note)

Users tend to use the same documentation type over time

H3: Doc. Type Predisposition

Usage in the previous three months
(Admin Documentation Usage
in prev. 3 months > 0)



Regression
Analysis

+ (3.31)

→ **Admin Doc. Time** > 0
(Pricing, Legal, Other)

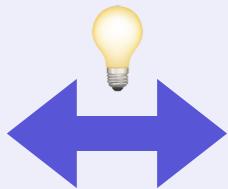


Documentation usage in page-view logs correlates with user characteristics

Platform Experience
Produce Experience
Web Search Intent
...



+ gender, roles, programming expertise, ...



Privacy-preserving Aggregated Logs

13:00 May 1, 2020

> Cloud SQL Overview

13:03 May 1, 2020

> Cloud SQL Pricing

13:05 May 1, 2020

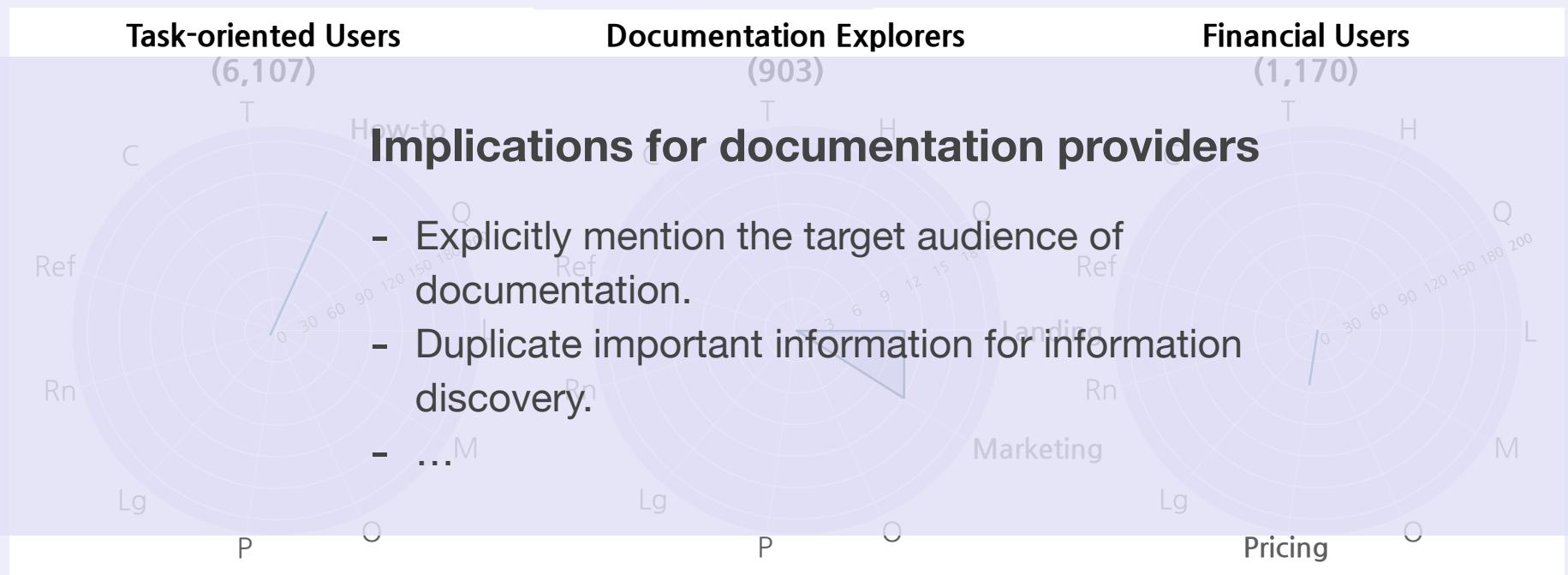
> Cloud SQL Key Terms

...



***Documentation log analysis allows
analyzing all documentation use,
with less effort and cost.***

Feasibility of using API documentation page-view logs to inform the design of software documentation pages





Documentation log analysis allows analyzing all documentation use, with less effort and cost.

Ask users
Survey & Interview

Robillard 09

Robillard 11

Uddin 15

Observe users
Lab/observation study

Meng 18

Sohan 17

NEW!
Analyze user logs

Can be cost-efficient

Can analyze the entire population

Can analyze actual usage

Understanding Documentation Use Through Log Analysis

An Exploratory Case Study of Four Cloud Services

Daye Nam, Andrew Macvean, Brad Myers, Bogdan Vasilescu



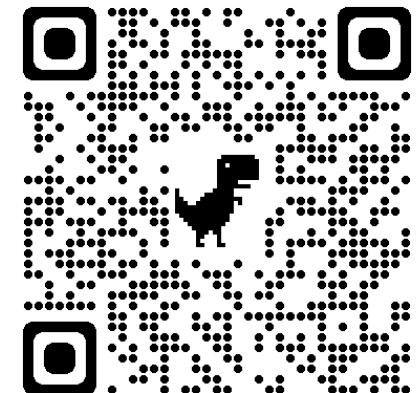
Documentation usage in page-view logs correlates with user characteristics and future product adoption



Documentation usage in page-view logs can be used in estimating users' needs and intent



Documentation log analysis allows analyzing all documentation use, with less effort and cost.



✉: dayen@cs.cmu.edu

🏡: <https://dayenam.com>



Carnegie Mellon University
School of Computer Science



Human-
Computer
Interaction
Institute

