

This document provides an overview of the files in the Teradata University Network Data Challenge 2016¹. The data set is provided by a non-profit organization - Hire Heroes USA². The organization is dedicated to the mission of empowering members of U.S. military, veterans and their spouses in their transition to the civilian workforce.

This document is supplemental to the main tutorial submission. The tutorial offers a hands-on exercise on ETL (Extract, Transform, Load) using Pentaho's Data Integration software.

1. *Data Statistics*

41 data files across 7 categories

The numbers in the parenthesis denote the number of files for each category

Facebook, LinkedIn, Twitter (20, 1, 1): Marketing analysis across each social media channel

Google analytics (10): How is the website doing in terms of number and source of visitors

Hubspot (1 file): Comparative picture of interactions and visits across social media channels

Salesforce (7 data files): Client Management and Employment Data

DonorPerfect (1 file): Fundraising analysis over time

It is important to note that while there is an overlap in the time field across the data sets above but they do not necessarily capture the data over the exact same timeline. For example, the data in the DonorPerfect file captures the data from July 2007 to January 2016 while the timeframe for the LinkedIn data set is from June 2014 to January 2016.

¹ <http://www.teradatauniversitynetwork.com/PARTNERS2016/2016-TUN-Data-Challenge/>

² <https://www.hireheroesusa.org/>

2. Data Definitions

2.1 DonorPerfect (The data is captured from July 2007 till January 2016)

Fieldname	Description
Date	Date on which donation was made
GL Code	Classifies the donation type across 11 categories such as Corporate, Individual, Grant etc.
Amount	Amount in dollars
Receipt Number	Optional field denoting a receipt number
Gift Type	Optional field denoting mode of payment - Check/Credit Card/Website/Wire Transfer
ID	Non-unique field

Sample analysis points:

- Trend in fundraising over time

2.2 LinkedIn (The data is captured from June 2014 till January 2016)

Fieldname	Description
DATE	Date of the post
company_update_clicks	The number of clicks on your content, company name, or logo.
company_update_unique_impressions	The number of times (unique) each update was shown to LinkedIn members.
distinct_pageview_overview_count	The number of views on company's page
company_update_impressions	The number of times each update was shown to LinkedIn members.
company_update_comments	The number of times people have commented on each update.
company_update_engagement	This percentage shows the number of interactions plus the number of clicks and followers acquired, divided by the number of impressions.

Overview of the Hire Heroes USA data sets

Submitted by: Dippy Aggarwal (Ph.D. Candidate), Karen C. Davis (Professor), University of Cincinnati, Ohio, USA

company_update_likes	The number of times people have liked each update.
company_update_shares	The number of times people have shared each update.
pageview_overview_count	The number of times people have viewed company's page
Source: https://help.linkedin.com/app/answers/detail/a_id/26032/~analytics-tab-for-company-pages	

Sample analysis

- Impressions by the day (Note: no time information available in the dataset)
- Engagement by the day

2.3 Twitter (The data is captured from September 2014 till January 2016)

You may notice missing descriptions for some of the fields in the schema descriptions in the document. Some of these fields names are self-explanatory (retweets/Tweet text etc.), some are having null for all data points in the dataset and there may be a few which require further exploration.

<i>Fieldname</i>	<i>Description</i>
Tweet id	
Tweet permalink	
Tweet text	
time	Date and time of the tweet
impressions	Times a user is served a Tweet in timeline or search results
engagements	Total number of times a user interacted with a Tweet. Clicks anywhere on the Tweet, including retweets, replies, follows, favorites, links, cards, hashtags, embedded media, username, profile photo, or Tweet expansion
engagement rate	Number of engagements divided by number of impressions
retweets	

Overview of the Hire Heroes USA data sets

Submitted by: Dippy Aggarwal (Ph.D. Candidate), Karen C. Davis (Professor), University of Cincinnati, Ohio, USA

replies	
likes	
user profile clicks	Clicks on the name, @handle, or profile photo of the Tweet author
url clicks	
hashtag clicks	
detail expands	Clicks on the Tweet to view more details
permalink clicks	
app opens	
app installs	
follows	Times a user followed you directly from the Tweet
email tweet	
dial phone	
media views	
media engagements	

Sample analysis points

- Impressions/Engagement rate by the time of the day
- Type of engagement: Retweets or Likes
- Which hashtags are performing the best?
- Any relationship between length of the tweet and engagement?

2.4 Hubspot (The data is captured from April 2015 till January 2016)

Field	Description
Status	SUCCESS
Channel	Facebook/Twitter/LinkedIn
Account	Hire Heroes USA
Campaign	Campaign type (Blank in some cases)
Published Message	Text for the post
Title	Null
Shortened Link	
Original Link	Link for the post
Clicks	Number of times user clicked on the post
Total Interactions	Sum of shares, likes and comments
Twitter Retweets	
Twitter Replies	
Twitter Favorites	
Facebook Likes	
Facebook Comments	
LinkedIn Likes	
LinkedIn Comments	
Publish Time	Date and time
Created Time	

Sample analysis points

- Engagement across channels
- How many registration campaigns launched over the last year by each channel?
- How many interactions and clicks did the campaign drove?

Overview of the Hire Heroes USA data sets

Submitted by: Dippy Aggarwal (Ph.D. Candidate), Karen C. Davis (Professor), University of Cincinnati, Ohio, USA

2.5 Facebook (The data is captured from July 2011 till January 2016)

Post Data (Maximum attributes: 51)

Post ID	
Permalink	
Post Message	
Type	Link/Photo/Status?
Countries	Null
Languages	Null
Posted	Date and time
Lifetime Post Total Reach	The total number of people your page post was served to. (Unique Users)
Lifetime Post organic reach	The number of people who saw your Page post in news feed or ticker, or on your Page's timeline. (Unique Users)
Lifetime Post Paid Reach	The number of people your advertised Page post was served to. (Unique Users)
Lifetime Post Total Impressions	
Lifetime Post Organic Impressions	The number of impressions of your post in News Feed or ticker or on your Page's Timeline. (Total Count)
Lifetime Post Paid Impressions	The number of impressions of your Page post in an Ad or Sponsored Story. (Total Count)
Lifetime Engaged Users	The number of people who clicked anywhere in your posts. (Unique Users)
Lifetime Talking About This (Post) by action type - comment	The number of unique people who created a story about your Page post by interacting with it. (Unique Users)
Lifetime Talking About This (Post) by action type - like	
Lifetime Talking About This (Post) by action type - share	
Lifetime Post Stories by action type - comment	The number of stories created about your Page post, by action type. (Total Count)
Lifetime Post Stories by action type - like	
Lifetime Post Stories by action type - share	
Lifetime Post Consumers	The number of people who clicked anywhere in your post. (Unique Users)

Overview of the Hire Heroes USA data sets

Submitted by: Dippy Aggarwal (Ph.D. Candidate), Karen C. Davis (Professor), University of Cincinnati, Ohio, USA

Lifetime Post Consumptions	The number of clicks anywhere in your post. (Total Count)
Lifetime Post consumers by type - link clicks	The number of people who clicked anywhere in your post, by type. (Unique Users)
Lifetime Post consumers by type - other clicks	
Lifetime Post consumers by type - photo view	
Lifetime Post Consumptions by type - link clicks	The number of clicks anywhere in your post, by type. (Total Count)
Lifetime Post Consumptions by type - other clicks	
Lifetime Post Consumptions by type - photo view	
Lifetime Negative feedback	The number of people who have given negative feedback to your post. (Unique Users)
Lifetime Negative Feedback from Users	The number of times people have given negative feedback to your post. (Total Count)
Lifetime Negative feedback - hide_all_clicks	The number of people who have given negative feedback to your post, by type. (Unique Users)
Lifetime Negative feedback - hide_clicks	
Lifetime Negative feedback - report_spam_clicks	
Lifetime Negative Feedback from Users by Type - hide_all_clicks	
Lifetime Negative Feedback from Users by Type - hide_clicks	
Lifetime Negative Feedback from Users by Type - report_spam_clicks	
Lifetime Post Impressions by people who have liked your Page	The number of impressions of your Page post to people who have liked your Page.(Total Count)
Lifetime Post reach by people who like your Page	The number of people who saw your Page post because they've liked your Page (Unique Users)
Lifetime Post Paid Impressions by people who have liked your Page	
Lifetime Paid reach of a post by people who like your Page	
Lifetime People who have liked your Page and engaged with your post	

Overview of the Hire Heroes USA data sets

Submitted by: Dippy Aggarwal (Ph.D. Candidate), Karen C. Davis (Professor), University of Cincinnati, Ohio, USA

Lifetime Average time video viewed	
Lifetime Organic views to 95%	Number of times your video was viewed to 95% of its length without any paid promotion. (Total Count)
Lifetime Organic views to 95%	Number of times your video was viewed to 95% of its length without any paid promotion. (Unique Users)
Lifetime Paid views to 95%	
Lifetime Paid views to 95%	
Lifetime Video length	Length of a video post (Total Count)
Lifetime Organic Video Views	Number of times your video was viewed for more than 3 seconds without any paid promotion. (Total Count)
Lifetime Organic Video Views	
Lifetime Paid Video Views	
Lifetime Paid Video Views	

Sample analysis

- Which type of posts is more popular – Link/Photo/Video?
- Consumption rate of videos vs photos
- Page fan engagement rate vs non-page fan engagement rate
- Negative feedback for each post type

Source: <https://searchenginewatch.com/sew/how-to/2206120/how-to-analyze-facebook-page-posts>

Overview of the Hire Heroes USA data sets

Submitted by: Dippy Aggarwal (Ph.D. Candidate), Karen C. Davis (Professor), University of Cincinnati, Ohio, USA

Facebook Page Data: Each of the data point is split across three dimensions: daily, weekly, and 28 days.

Date
Lifetime Total Likes
Daily New Likes
Daily Unlikes
Daily Page Engaged Users
Daily/weekly/28 days Like Sources - ads and several other categories
Total/Daily/Weekly/28 days Reach - organic, paid
Likes by logged-in users
Total/Daily/Weekly Impressions - organic/paid
Daily/weekly/28 days frequency distribution
People Talking about this by story type
Total Consumers
Total Consumptions
Daily/weekly/28 days - People who interacted with your page by content type
Page consumptions by type
Lifetime likes by gender and age
Lifetime likes by country
Lifetimes likes by city
Lifetime likes by language
Weekly Reach by City
Weekly Reach by Country
Daily Country: People Talking about this
Daily City: People Talking about this
Daily External Referrers
Negative feedback
Positive feedback
Checkins
Video views
Daily Number of time a video has seen again for 30s or to the end

2.6 Google Analytics

Filename	Number of rows	Number of attributes
Analytics Hire Heroes Audience Overview 20120401-20160106	1376	2
Analytics Hire Heroes Browser _ OS 20120401-20160106	10	10
Analytics Hire Heroes Desktop vs Mobile 20120401-20160106	3	5
Analytics Hire Heroes Engagement - Page Depth 20120401-20160106	20	3
Analytics Hire Heroes Engagement - Session Duration 20120401-20160106	7	3
Analytics Hire Heroes Frequency _ Recency 20120401-20160106	14	3
Analytics Hire Heroes New vs Returning 20120401-20160106	2	7 out of 10 containing non-null/non-zero values
Analytics Hire Heroes Pages - Explorer 20120401-20160106	249	7 out of 8 contain non-null or non-zero values
Analytics Hire Heroes Pages - Navigation Summary 20120401-20160106	503	2
Analytics Hire Heroes Site Speed Page Timings 20120401-20160106	1479	5 out of 6

Overview of the Hire Heroes USA data sets

Submitted by: Dippy Aggarwal (Ph.D. Candidate), Karen C. Davis (Professor), University of Cincinnati, Ohio, USA

Analytics Audience

Day Index

Sessions

Analytics Browser

Browser

Sessions

% New Sessions

New Users

Bounce Rate

Pages / Session

Avg. Session Duration

Goal Conversion Rate

Goal Completions

Goal Value

Day Index

Sessions

Overview of the Hire Heroes USA data sets

Submitted by: Dippy Aggarwal (Ph.D. Candidate), Karen C. Davis (Professor), University of Cincinnati, Ohio, USA

Device Category
Total Sessions
Desktop Sessions
Mobile Sessions
Tablet Sessions

Analytics Desktop vs Mobile 20120401-20160106

Page Depth
Sessions
Pageviews

Analytics Engagement - Page Depth 20120401-20160106

Session Duration
Sessions
Pageviews

Analytics Engagement - Session Duration 20120401-20160106

Count of Sessions
Sessions
Pageviews

Analytics Frequency _ Recency 20120401-20160106

User Type
Sessions
% New Sessions
New Users
Bounce Rate
Pages / Session
Avg. Session Duration
Goal Conversion Rate
Goal Completions
Goal Value

Day Index
Sessions

Hire Heroes New vs Returning 20120401-20160106

Overview of the Hire Heroes USA data sets

Submitted by: Dippy Aggarwal (Ph.D. Candidate), Karen C. Davis (Professor), University of Cincinnati, Ohio, USA

Page		
Pageviews		
Unique Pageviews		
Avg. Time on Page		
Entrances		
Bounce Rate		
% Exit	Day Index	
Page Value	Pageviews	Analytics Hire Heroes Pages - Explorer 20120401-20160106

Previous Page Path	
Pageviews	Analytics Hire Heroes Pages - Navigation Summary 20120401-20160106

Browser		
Avg. Page Load Time (sec)		
Pageviews		
Bounce Rate		
% Exit	Day Index	
Page Value	Avg. Page Load Time (sec)	Analytics Hire Heroes Site Speed Page Timings 20120401-20160106

2.7 Salesforce Data

The dataset contains both the actual data and schema description files.