The DHI Vacancy and Application Flow Database: Record Layouts, Variable Descriptions, and Summary Statistics

Steven J. Davis and Brenda Samaniego de la Parra*

2 February 2019

Overview

The ***DHI Vacancy and Application Flow Database*** (hereafter, "DHI Database") contains information related to job vacancies posted by DHI clients from January 2012 to December 2017. The DHI Database has four main files: ***Accounts, Jobs, Activity***, and ***Detailed Applications***. ***Accounts*** contains information about each client that posts one or more jobs such as its industry and billing address. ***Jobs*** contains information about each job posting, such as job description and intended place of employment. ***Activity*** contains daily information for each day each posting is online, such as the number of views per day, the daily number of applications completed via email, and the number of applicants who click through to an external application system for each day a job is active on the DHI platform. ***Detailed Applications*** contains information regarding each application including the date and time, down to the second, in which it was received and the applicant's current job title and location. ***Detailed Applications*** also includes an applicant identifier that allows grouping all the applications submitted by a particular job seeker. Each observation in ***Jobs*** can be merged to data about the employer posting the job in ***Accounts,*** daily activity information in ***Activity,*** and information about the applicant in ***Detailed Applications,*** allowing for job-level, firm-level, regional, industry, and applicant-level analyses.

The rest of this technical memorandum provides additional information about the ***Accounts, Jobs***, ***Activity,*** and ***Detailed Applications*** files: Record layouts, variable names and descriptions, and basic summary statistics. We report the number of distinct values for character/text variables is reported after basic data cleaning (removing punctuation, lower case letters, duplicate spacing).

The *Accounts* File

File name: accounts.sas7bdat
File size: 12.3 MB
Encoding: wlatin1 Western (Windows)
Total Number of Observations: 60,628
Total Number of Variables: 14

The Accounts File records information about clients that post vacancies on the DHI website.

Variable Order, Definitions, Admissible Values, and Summary Statistics:

1. **account_id_anom** – Automatically generated 4 to 7-digit numeric identifier for each observation in the Accounts File. Account_id uniquely identifies each client and its different branches, geographical locations or business units. 22% of the clients registered with DHI at the end of 2016 had not yet posted any jobs. Variable type: Numeric; Variable Length: 7; Number of observations with missing values: 0; Number of unique non-missing values: 60,628.

2. **account_num_anom** – A 9-character alpha-numeric identifier for client accounts. A company can have more than one account_num, and an account_num can be associated with more than one account_id. For example, a company can open various accounts with DHI for each of its locations or business units. Each of these locations or units will have a different account id but the same account number. The company can also choose to have a different account for each (or some) of these locations or units. Variable type: Character; Variable Length: 9; Number of observations with missing values: 0; Number of unique non-missing values: 57,407.

3. **company_name_anom** – Company name's were anonymized into a "Company in City X" format to maintain DHI clients' confidentiality. The number of unique company names is less than the number of unique account_num values because a company can have more than one contractual relationship with DHI, for example through different geographical locations, business units, etc. Variable type: Character; Variable Length: 60; Number of observations with missing values: 10; Number of unique non-missing values before anonymization: 54,399.

Account_Num Distribution by No. of Distinct Account_ID's				
Number of Distinct Account ID (account_id) per Account Number (account_num)	No. of Accounts (identified using account_num)	% of Accounts (identified using account_num)	Cumulative Frequency	Cumulative Percent
1	55,204	96.2	55,204	96.2
2	1,932	3.4	57,136	99.5
3	153	0.3	57,289	99.8
4	32	0.1	57,321	99.9
5	23	0.0	57,344	99.9
6	17	0.0	57,361	99.9
7	6	0.0	57,367	99.9
8	4	0.0	57,371	99.9
9	9	0.0	57,380	100.0
10	5	0.0	57,385	100.0
11	2	0.0	57,387	100.0
12	2	0.0	57,389	100.0
13	1	0.0	57,390	100.0
14	1	0.0	57,391	100.0
15	4	0.0	57,395	100.0
16	2	0.0	57,397	100.0
18	2	0.0	57,399	100.0
20	1	0.0	57,400	100.0
23	1	0.0	57,401	100.0
24	1	0.0	57,402	100.0
28	1	0.0	57,403	100.0
32	1	0.0	57,404	100.0
43	1	0.0	57,405	100.0
59	1	0.0	57,406	100.0
121	1	0.0	57,407	100.0

4. **company_type** – Indicates the type of organization posting job vacancies. Variable type: Character; Variable Length: 16; Number of observations with missing values: 0; Number of unique non-missing values: 5.

   o Direct Hire = Companies who post vacancies to hire employees of their own account. These companies seek to augment their own labor force.

   o Recruitment/Staffing Firm = Companies that post vacancies to find workers who will provide their services to other organizations or firms. This includes both:

      a. companies that hire employees to provide their services at a different organization under, for example, an outsourcing arrangement, and

      b. companies that post vacancies on behalf of a different organization. They conduct the recruitment efforts on behalf of another firm who is then responsible for making the final hiring decision.

   o RPO = Recruitment Process Outsource. A company that outsources the hiring of its workforce. Unlike a recruitment firm, under an RPO arrangement:

      a. the recruitment team works from the client's location, as opposed to from their own offices, and

      b. the additional recruitment efforts provided by the RPO are paid for regardless of placements achieved, unlike a typical recruitment firm arrangement.

   o Ad/Media Agency = Job posting and worker placement companies that match workers and firms with vacancies.

   o Unassigned

Account ID Distribution by Company Type		
Company Type	**No. of Accounts**	**Share of Accounts**
Direct Hire	47,951	79.1
Recruitment Firm	10,463	17.3
Unassigned	1,883	3.1
Ad/Media Agency	275	0.5
RPO	56	0.1

6. **company_city** – Name of the city on the company's billing address when it established its DHI account. Company_city may differ from the intended place of employment for the

company's job postings. Variable type: Character; Variable Length: 17; Number of observations with missing values: 556; Number of unique non-missing values: 4,653

7. **company_state** – State or province on the company's billing address when it established its DHI account. Company_state may differ from the intended place of employment for the company's job postings. For US companies, company_state is the 2-character abbreviation of the state. For Non-US companies, the state may or may not be abbreviated. Variable type: Character; Variable length: 4; Number of observations with missing values: 797; Number of unique non-missing values: 124.

Account ID Distribution by Company State (US Companies)
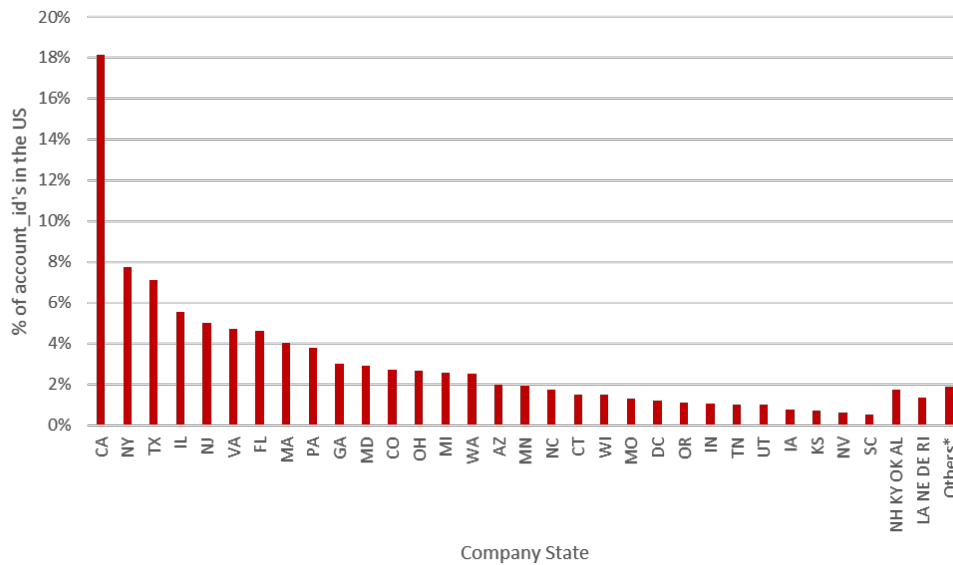
Notes: Others includes AR, NM, ID, ME, MT, VT, AK, WV, MS, HI, WY, SD, and ND.

8. **company_country** – Country on the company's billing address when it established its DHI account. Variable type: Character; Variable Length: 15; Number of observations with missing values: 266; Number of unique non-missing values: 57.

Account Distribution by Country		
Country	Number of Accounts	Share of Accounts
United States	59,620	98.8
Canada	300	0.3
India	162	0.2

|		
United Kingdom	153	0.2
Australia	19	0.0
Germany	9	0.0
Switzerland	7	0.0
Philippines	6	0.0
Netherlands	5	0.0
Others*	85	0.0

Notes: Others includes New Zealand, Cayman Islands, France, Ireland, United Arab Emirates, Bermuda, Hong Kong, Israel, Saudi Arabia, Belgium, Bahamas, China, Malaysia, Mexico, Singapore, Virgin Islands, Argentina, Czech Republic, Indonesia, Italy, Japan, Luxembourg, Malta, Pakistan, Qatar, South Africa, Spain, Sweden, and Vietnam.

9. **dice_metro** – The metropolitan statistical area (MSA) corresponding to the company's billing address. The value corresponds to the MSAs as defined by the Office of Management and Budget and assigned to an account based on the clients' billing address proximity to the MSA. DHI assigns an MSA to an account if its billing address is within 30 miles of the MSA's centroid. 352 accounts in the US are unassigned to a MSA (value of dice_metro equals "NONE"). Variable type: Character; Variable Length: 51; Number of observations with missing values: 90; Number of unique non-missing values: 724.

10. **annual_revenue_range** – Annual company revenue range in US dollars. Revenue data is from D&B. These revenue values and the other D&B data do not necessarily pertain to the most recent calendar or fiscal year. Instead, it likely reflects the annual revenue at the time the account with DHI was first established. 6,986 accounts have annual revenues equal to 0. Variable type: Numeric. Number of observations with missing values: 9,320.

Annual Revenue Percentiles (rounded)	
Minimum	$0
1st Pctl	$0
10th Pctl	$0
25th Pctl	$240,000
50th Pctl	$4,000,000
75th Pctl	$35,000,000

90th Pctl	$217,000,000
99th Pctl	$5,550,000,000
Maximum	$94,600,000,000
Mean	$294,000,000
Std Dev	$2,500,000,000

Accounts' Log of Annual Revenue Histogram
(for accounts with positive revenues only)
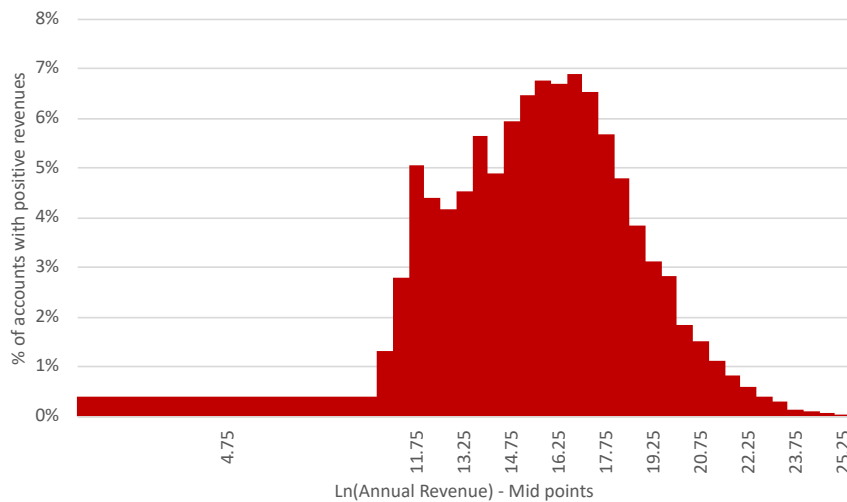
11. industry – Industry category for the company's line of operations, as reported by D&B. Variable type: Character; Variable Length: 16; Number of observations with missing values: 9,300; Number of unique non-missing values: 38.

Account Distribution by Industry					
Industry	**No. of Accounts**	**Share of Accounts**	**Industry**	**No. of Accounts**	**Share of Accounts**
Consulting	10,793	21.0	Other	8,357	16.3
Technology	7,402	14.4	Retail	4,072	7.9
Manufacturing	2,525	4.9	Education	2,467	4.8
Healthcare	1,823	3.6	Finance	1,333	2.6
Electronics	1,273	2.5	Government	1,240	2.4
Banking	1,104	2.2	Telecommunications	1,052	2.0

Not For Profit	1,019	2.0	Insurance	1,007	2.0
Communications	767	1.5	Engineering	747	1.5
Transportation	617	1.2	Entertainment	540	1.1
Construction	484	0.9	Utilities	400	0.8
Machinery	385	0.8	Chemicals	376	0.7
Apparel	316	0.6	Hospitality	258	0.5
Biotechnology	242	0.5	Food & Beverage	163	0.3
Media	155	0.3	Agriculture	133	0.3
Energy	117	0.2	Recreation	101	0.2
Shipping	50	0.1	Recruitment Generic	2	0.0
Biopharmaceutical	2	0.0	Information Technology	2	0.0
Recruitment Core	1	0.0	Ad Agency	1	0.0
General	1	0.0	Media/Agency	1	0.0

12. **ownership** – The company's ownership structure, as reported by D&B. Variable type: Character; Variable Length: 10; Number of observations with missing values: 9,349; Number of unique non-missing values: 5.

o Private = Privately held companies.

o Public = Privately owned companies listed on a public stock exchange.

o Government = Public sector, including government-owned companies.

o Subsidiary

o Other = Other organizations, e.g., NGOs.

| Account Distribution by Ownership Structure		
Ownership	No. of Accounts	Share of Accounts
Private	48,443	94.5
Public	1,584	3.1
Government	1,228	2.4
Other	23	0.0
Subsidiary	1	0.0

13. num_of_employees_category – Range for the number of employees at the company. D&B is the source for this variable. There are 5,652 accounts with 0 employees. Variable type: Numeric; Number of observations with missing values: 9,310.

Number of Employees Percentiles (rounded)	
Minimum	0
1st Pctl	0
10th Pctl	0
25th Pctl	4
50th Pctl	30
75th Pctl	200
90th Pctl	1,000
99th Pctl	23,000
Maximum	434,000
Mean	1,200
Std Dev	9,600

Account Distribution by Number of Employees
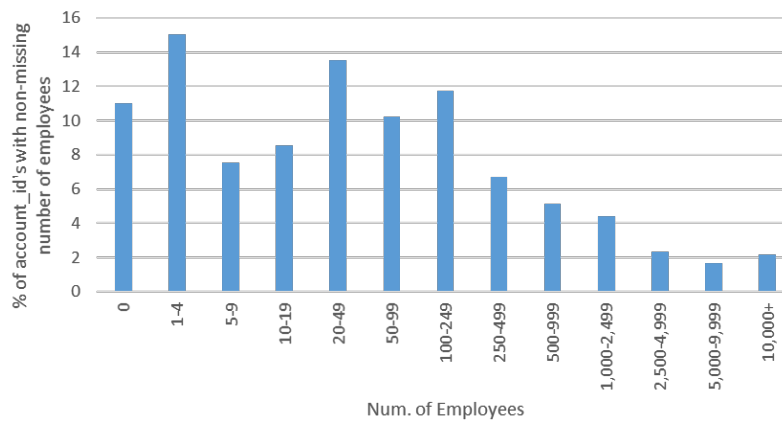
Weighted Account Distribution by Number of Employees
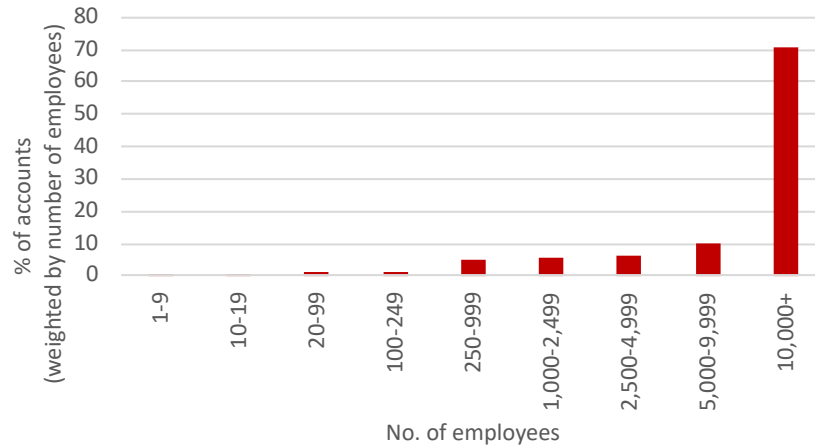
14. **sic** – The 4-digit Standard Industrial Classification code (SIC 2007) for the company's line of operations, as reported by D&B. Variable type: Numeric; Number of observations with missing values: 9,967; Number of unique non-missing values: 885.

15. **sic_desc** – Description of the 4-digit Standard Industrial Classification description, as reported by D&B. Variable type: Character; Variable Length: 48; Number of observations with missing values: 10,030; Number of unique non-missing values: 889.

	Account Distribution by Industry (Top 40 Industries)						
SIC4	Description	Number of Accounts	Share of Accounts	SIC4	Description	Number of Accounts	Share of Accounts
7371	Custom Computer Programming Services	4,820	9.5	7379	Computer Related Services, Nec	3,583	7.1
8742	Management Consulting Services	2,239	4.4	8748	Business Consulting, Nec	2,014	4.0
7361	Employment Agencies	1,903	3.8	7389	Business Services, Nec	1,666	3.3
9999	Nonclassifiable Establishments	1,616	3.2	7373	Computer Integrated Systems Design	1,360	2.7
7372	Prepackaged Software	1,229	2.4	8221	Colleges and Universities	987	2.0
7363	Help Supply Services	687	1.4	8711	Engineering Services	654	1.3

8741	Management Services	617	1.2	5045	Computers, Peripherals, and Software	588	1.2
4813	Telephone Communication, Except Radio	576	1.1	7374	Data Processing and Preparation	572	1.1
8211	Elementary and Secondary Schools	523	1.0	7311	Advertising Agencies	516	1.0
8011	Offices and Clinics of Medical Doctors	497	1.0	8111	Legal Services	492	1.0
9111	Executive Offices	489	1.0	6411	Insurance Agents, Brokers, and Service	453	0.9
5734	Computer and Software Stores	380	0.8	8731	Commercial Physical Research	370	0.7
8621	Professional Organizations	346	0.7	8062	General Medical and Surgical Hospitals	338	0.7
8299	Schools and Educational Services	325	0.6	8099	Health and Allied Services, Nec	308	0.6
8322	Individual and Family Services	288	0.6	6531	Real Estate Agents and Managers	279	0.6
8721	Accounting, Auditing, and Bookkeeping	266	0.5	8732	Commercial Nonphysical Research	249	0.5
8999	Services, Nec	237	0.5	7375	Information Retrieval Services	235	0.5
8733	Noncommercial Research Organizations	228	0.5	5961	Catalog and Mail-order Houses	226	0.4
4899	Communication Services, Nec	218	0.4	8222	Junior Colleges	217	0.4
6282	Investment Advice	216	0.4	8611	Business Associations	211	0.4

Notes: Nec = Not elsewhere classified.

The *Jobs* File

File Name: jobs.sas7bdat
File size: 7.23 GB
Encoding: latin1 Western (ISO)
Total Number of Observations: 7,610,456
Total Number of Variables: 12

The Jobs file contains information about the individual jobs posted by DHI clients.

Variable Order, Definitions, Admissible Values, and Summary Statistics:

1. **job_id** – A unique alpha-numeric identifier for each job posting. The *Activity* and *Detailed Applications* Files also contains this variable which allows merging the datasets to track the postings' online spells and applications received through time. Variable type: Character; Variable Length: 32; Number of observations with missing values: 0; Number of unique non-missing values: 7,610,456.

Job Postings by First Active Date
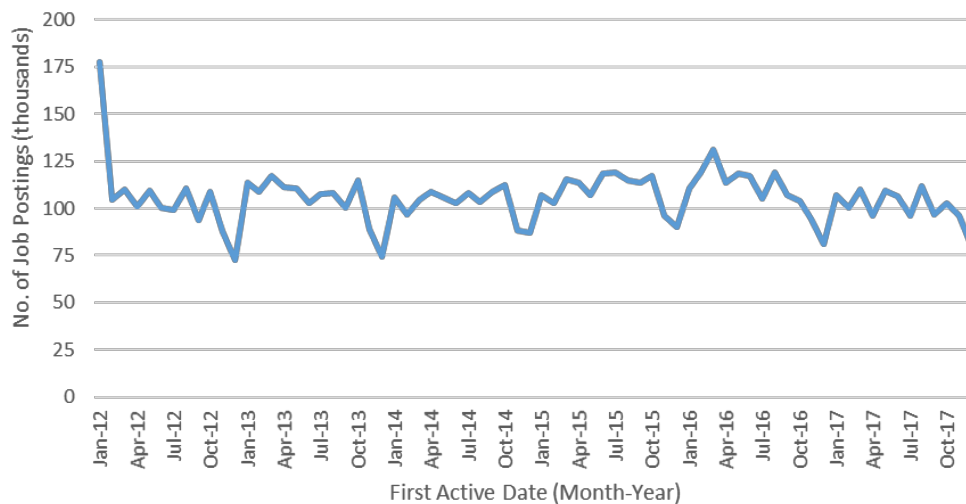
2. **account_id_anom** – Automatically generated 4 to 7-digit numeric identifier for each client account. Account_id can be used to merge the *Accounts* and *Jobs* files. Variable type: Numeric; Variable Length: 7; Number of observations with missing values: 0; Number of unique non-missing values: 47,346.

3. **account_num_anom** – A 9-character alpha-numeric identifier for client accounts. A company can have more than one account_num, and an account_num can be associated with more than one account_id. See additional details in the *Accounts File* description. Variable type: Character; Variable Length: 9; Number of observations with missing values: 0; Number of unique non-missing values: 44,583.

4. **job_title** – Job title is an open text field where the company creating the posting can write a description of the job. It can include a combination of the job's function or occupation, required skills, and seniority level. For some jobs, this variable also includes a brief description of the company or the job characteristics (e.g. "Great opportunity!", "Work-life balance", "Innovative start-up"). Variable type: Character; Variable Length: 247; Number of observations with missing values: 205; Number of unique non-missing values: 2,473,274.

Common Job Functions Mentioned in job_title					
Job Function	**No. of Postings**	**Share**	**Job Function**	**No. of Postings**	**Share**
DEVELOPER	1,785,520	23.5%	ENGINEER	1,189,417	15.6%
ANALYST	968,355	12.7%	MANAGER	724,112	9.5%
ARCHITECT	438,111	5.8%	CONSULTANT	409,223	5.4%
PROJECT MANAGER	363,591	4.8%	ADMINISTRATOR	353,990	4.7%
BUSINESS ANALYST	280,391	3.7%	SPECIALIST	183,953	2.4%
TECHNICIAN	135,584	1.8%	SALES	121,697	1.6%
TESTER	118,351	1.6%	PROGRAMMER	115,200	1.5%
DESIGNER	94,394	1.2%	HELP	86,002	1.1%
DIRECTOR	67,462	0.9%	COORDINATOR	58,684	0.8%
PROGRAM MANAGER	47,140	0.6%	BUSINESS SYSTEMS ANALYST	42,358	0.6%
SYSTEM ADMINISTRATOR	38,973	0.5%	EXPERT	37,541	0.5%
BUSINESS INTELLIGENCE	33,207	0.4%	SCRUMMASTER	25,378	0.3%
MODELER	23,034	0.3%	RECRUITER	20,849	0.3%
SCIENTIST	15,929	0.2%	INTERN	15,615	0.2%
ADVISOR	14,289	0.2%	SUPERVISOR	9,742	0.1%

BUSINESS DEVELOPMENT	9,728	0.1%	CUSTOMER SERVICE	9,083	0.1%
AUDITOR	8,107	0.1%	ELECTRICAL ENGINEER	7,315	0.1%
COACH	7,009	0.1%	MECHANICAL ENGINEER	6,883	0.1%
ACCOUNTANT	6,424	0.1%	ADMINISTRATIVE ASSISTANT	6,100	0.1%
REPORT WRITER	4,529	0.1%	PRESIDENT	4,290	0.1%
CLERK	3,765	0.0%	QUALITY ASSURANCE TESTER	3,750	0.0%
RESEARCHER	3,695	0.0%	NURSE	2,679	0.0%
GENERALIST	2,101	0.0%	STATISTICIAN	1,852	0.0%
INSPECTOR	1,590	0.0%	WEBMASTER	1,045	0.0%
INFORMATION OFFICER	727	0.0%	CHEMIST	713	0.0%
SYSTEM INTEGRATOR	664	0.0%	TECHNOLOGY OFFICER	512	0.0%
DEVELOPMENT OPERATIONS ENGINEER	437	0.0%	NINJA	402	0.0%
CENTER OPERATOR	400	0.0%	INCIDENT RESPONDER	207	0.0%

Notes: Shares do not add to 100%. Postings can mention more than one function in their job titles and the list of job functions is not exhaustive.

5. **product_type** – Describes the type of contract used by the client to post a job on DHI's website. A client can have different contracts for each job posting. This variable is missing for jobs posted after October 2017. Variable type: Character; Variable Length: 21; Number of observations with missing values: 214,381; Number of unique non-missing values: 8.

   o Subscription: This contract type grants the client a set number of "job slots" that are available for posting jobs. The client can freely allocate (and reallocate) job postings across slots, but the number of postings cannot exceed the number of slots at a point in time.

   o Classified: This contract type grants the client a set number of "job credits" that are available for posting jobs. For each credit, the client can post a single job. Unlike job slots, job credits cannot be reused for a different posting.

o Webstore: These are jobs posted directly by the client in a "self-service" mode. Clients can choose the number of job credits they want among pre-defined set. Each credit can be used to post a single job for 30 days at any time within the following 12 months.

o Subscription+PPV: This contract type appends a pay-for-performance clause to the subscription contract described above. The pay-per-performance clause specifies an upper limit on the total number of applicant views accumulated by all job postings under the contract per month. Upon reaching the limit, all the postings are taken offline and are no longer visible to applicants.

o One Post: DHI enters on-going agreements with recruiting agencies and staffing firms to handle single job postings on a long-term basis without requiring a signature/contract for every job post. This agreement helps streamline the job posting function for customers with frequent posting needs. The customer can purchase single 30-day job credits as needed at any time under this long-term contract.

o Classified+PPV: This contract type appends a pay-for-performance clause to the classified contract described above. The pay-per-performance clause specifies an upper limit on the total number of applicant views accumulated by all job postings under the contract per month. Upon reaching the limit, all the postings are taken offline and are no longer visible to applicants.

o Unspecified or None: The contract type is not specified in the database.

Job Posting Distribution by Product_Type at the Time of Posting		
Product Type	**Number of Postings**	**Share of Postings**
Subscription	7,236,342	97.8%
Webstore	60,887	0.8%
Classified	54,124	0.7%
Subscription + PPV	34,166	0.5%
One Post	10,493	0.1%
Classified + PPV	48	0.0%
Unspecified	14	0.0%
none	1	0.0%

Notes: Shares are calculated as a percent of postings with non-missing contract type.

As described in the "product_type" variable, DHI contracts are associated with a certain number of job slots, or job credits. Furthermore, contracts with a performance clause have a limit on total views by potential applicants. The following 3 variables refer to each contracts' job slots, job credits, or viewing limits, if any.

6. **job_slots** – Number of job slots available for job postings under a Subscription contract. For postings starting in Variable type: Numeric; Number of observations with missing values: 328,523. None of the job postings where product_type is "Subscription" or "Subscription + PPV" have missing values in the job_slots variable. This variable is either missing or equal to 5 for jobs posted after October 2017. Number of unique non-missing values: 276.

Job Slots (for Job Postings with Subscription and Subscription + PPV Product Types)		
	Jan. 2012 – Oct. 2017	**Nov.-Dec. 2017**
Minimum	0	5
1st Pctl	2	5
10th Pctl	5	5
25th Pctl	5	5
50th Pctl	17	5
75th Pctl	120	5
90th Pctl	1,000	5
99th Pctl	11,619	5
Maximum	20,008	5
Mean	602.48	5
Std Dev	2,327.74	0
N	6,266,722	1,003,786

Notes: The table excludes postings with non-Subscription product type.

7. **job_credits** – Number of job credits available for vacancy postings under a Classified contract. Variable type: Numeric; Number of observations with missing values: 7,564,086.

Most of these missing values are expected, since they appear on "Subscription" type postings which do not have job credits. This variable is missing for jobs posted after October 2017. Number of unique non-missing values: 32.

Job Credits (for Job Postings with non-Subscription Product Types)	
	Jan. 2012 – Oct. 2017
Minimum	1
1st Pctl	1
10th Pctl	1
25th Pctl	1
50th Pctl	1
75th Pctl	5
90th Pctl	13
99th Pctl	50
Maximum	80
Mean	4.81
Std Dev	9.26
N	46,215

Notes: The table excludes job postings with "Subscription" and "Subscription + PPV" jobs. It also excludes non-subscription jobs with missing values for job_credits.

8. **ppv_job_view_limit** – The upper limit on monthly views specified in the pay-for-performance contract. Variable type: Numeric; Number of observations with missing values: 7,576,239. This variable is missing for jobs posted after October 2017. None of the jobs posted before November 2017 where product_type includes a performance clause ("PPV") have missing values in the ppv_job_view_limit variable. Number of unique non-missing values: 27.

Job Views Limit Performance Clause (for Job Postings with Performance Clause Contracts)	
	Jan. 2012 – Oct. 2017
Minimum	1
1st Pctl	1
10th Pctl	15

25th Pctl	15
50th Pctl	15
75th Pctl	23
90th Pctl	90
99th Pctl	99
Maximum	99
Mean	26.96
Std Dev	24.71
N	34,214

Notes: The table includes postings with performance clause contracts ("PPV") with non-missing values for ppv_job_view_limit.

9. **first_active_date** – Date the job is first posted in year-month-day (e.g. 2017-05-22) format. Jobs posted on the DHI website can be taken offline and reposted multiple times. When offline, the posting is not visible to potential applicants. First_active_date refers to the date a particular job_id was ever visible to applicants, therefore, the value of this variable does not change whenever there are offline periods or re-postings. Variable type: Date. Number of observations with missing values: 0; Number of unique non-missing values: 2,192.

Job Postings by First Active Date Year		
First Active Year	**Number of Postings**	**Percent**
2012	1,274,429	16.7
2013	1,258,298	16.5
2014	1,231,310	16.2
2015	1,313,767	17.3
2016	1,319,847	17.3
2017	1,212,805	15.9

10. **last_active_date** – Last date that the job_id was online. The format is year-month-day (e.g. 2017-05-22). Variable type: Date. Number of observations with missing values: 0; Number of unique non-missing values: 2,210.

Job Postings by Last Active Date Year		
Last Active Year	No. of Postings	Percent
2012	1,168,929	15.4
2013	1,255,450	16.5
2014	1,230,247	16.2
2015	1,316,198	17.3
2016	1,335,154	17.5
2017	1,304,478	16.2
2018	70,531	0.9

11. **job_city** – Intended city of employment, as specified by the company posting the job. Variable type: Character; Variable Length: 30; Number of observations with missing values: 8,131; Number of unique non-missing values: 21,155.

12. **job_state** – Intended state of employment, as specified by the company posting the job. This variable is often missing for jobs outside of the US. Variable type: Character; Variable Length: 10; Number of observations with missing values: 40,479; Number of unique non-missing values: 374.

13. **tax_terms** – Terms of employment specified by the company posting the job. Variable type: Character; Variable Length: 25; Number of observations with missing values: 255,861; Number of unique non-missing values: 290. Tax terms can be any combination of the following 8 items:

o FULLTIME = Full-time position

o PARTTIME = Part-time position

o CON_CORP = Corporation-to-corporation contract (3rd party). The job poster is willing to accept job applications from 3rd party vendors or "middleman". The applicant will work on a contract through the vendor in a business-to-business type of arrangement.

o CON_HIRE_CORP = Corporation-to-corporation contract-to-hire (3rd party). This is a corporation-to-corporation contract with a specific clause enabling the employer to hire the person after he has worked for the company for some time.

- CON_IND = Contract on 1099 (1st party). The person hired is considered self-employed (filling a 1099 tax form) and represents herself with the job poster. The job is normally conducted under an hourly contract.
- CON_HIRE_IND = Contract-to-hire on 1099 (1st party). This is a 1099 contract with a specific clause enabling the employer to hire the person after he has worked for the company for some time.
- CON_W2 = Contract on W2 (1st party). Under this contract, the person is hired as an employee (files W2 tax forms) and represents herself with the job poster.
- CON_HIRE_W2 = Contract-to-hire on W2 (1st party). This is a W2 contract with a specific clause enabling the employer to hire the person after he has worked for the company for some time.

Job Postings by Tax Terms					
Tax Term	No. of Postings	%	Tax Term	No. of Postings	%
CONW2	3,932,838	23.2%	FULLTIME	3,387,045	20.0%
CONCORP	2,726,217	16.1%	CONIND	2,569,196	15.2%
CONHIREW2	1,561,947	9.2%	CONHIREIND	1,246,194	7.4%
CONHIRECORP	1,178,659	7.0%	PARTTIME	314,990	1.9%

Note: Shares add to more than 100% because a posting can, and often does, include more than one tax terms.

14. **telecommute_option** – Yes/No variable indicating whether the job accepts telecommuting. Variable type: Character; Variable Length: 3; Number of observations with missing values: 0; Number of unique non-missing values: 3.

Job Postings by Telecommuting Options		
Telecommute Option	No. of Postings	Percent
No	7,557,992	99.3
Yes	41,875	0.6
N/A	10,589	0.1

15. **pay_rate** – Free-text field in which clients can indicate compensation terms. 5.5% of job postings indicate a $0 pay. It is common for clients to indicate that pay rate is "Negotiable",

"Competitive", or "Market Rate". Variable type: Character; Variable Length: 100; Number of observations with missing values: 1,408,676.

| Top 20 Pay Rates (per Year and per Hour) | | | | | |
| Pay per Year | | | Pay per hour | | |
Pay Rate	No. of Postings	%	Pay Rate	No. of Postings	%
$100,000	61,312	4.8%	$50	76,806	6.0%
$90,000	51,154	4.0%	$60	60,153	4.7%
$80,000	50,868	4.0%	$55	51,281	4.0%
$120,000	32,095	2.5%	$40	45,518	3.6%
$70,000	31,773	2.5%	$45	44,142	3.5%
$85,000	30,014	2.4%	$70	34,712	2.7%
$75,000	29,248	2.3%	$65	33,420	2.6%
$110,000	28,766	2.3%	$35	24,640	1.9%
$60,000	25,032	2.0%	$30	23,027	1.8%
$65,000	18,506	1.4%	$80	21,574	1.7%
$95,000	18,393	1.4%	$75	19,679	1.5%
$50,000	17,354	1.4%	$25	17,344	1.4%
$130,000	16,041	1.3%	$20	16,756	1.3%
$125,000	12,358	1.0%	$15	12,580	1.0%
$150,000	12,214	1.0%	$90	11,955	0.9%
$55,000	11,747	0.9%	$100	10,543	0.8%
$115,000	11,041	0.9%	$18	8,992	0.7%
$140,000	10,311	0.8%	$85	8,671	0.7%
$40,000	8,405	0.7%	$48	6,096	0.5%
$45,000	7,900	0.6%	$17	5,982	0.5%

Note: Percent values are calculated as shares from 1,359,166 postings with valid numeric values for pay_rate. For job postings that specify a range, the table above considers only the lower end of a range.

The *Activity* File

File Name: activity.sas7bdat
File Size: 38.1 GB
Encoding: latin1 Western (ISO)
Total Number of Observations: 179,574,177
Total Number of Variables: 21

Each observation in the *Activity* File is a job_idXdate combination. There is one observation for every day when a job posting is active (i.e. each observation is a vacancy day). A job posting is considered active on a given day if it is visible to applicants for any amount of time during that day. Among other things, the *Activity* File allows calculating offline spells for each posting by identifying non-sequential observations.

When posting a job, the client decides whether applicants will submit applications via email through DHI or through an external application system operated by the client or by a third party. In the first case, the *Activity* File records the number of completed applications via email during the day. In the second case, it records the day's number of viewers redirected to the external application site (URL). Postings cannot receive both types of applications simultaneously, but the client can choose to switch the application channel at any point throughout the life of a posting.

Variable Order, Definitions, Admissible Values, and Summary Statistics:

1. **job_id** – A unique alpha-numeric identifier for each job posting. This variable can be used to merge the *Jobs, Activity,* and *Detailed* Files. Variable type: Character; Variable Length: 32; Number of observations with missing: 0; Number of unique non-missing values: 7,838,006.

2. **account_id_anom** – Automatically generated 4 to 7-digit numeric identifier for each client account. Account_id can be used to merge the *Accounts*, *Jobs,* and *Activity* Files. See the *Accounts* File description for additional information on this variable. Variable type: Numeric; Variable Length: 7; Number of observations with missing values: 0; Number of unique non-missing values: 60,628.

3. **account_num_anom** – A 9-character alpha-numeric identifier for client accounts. A company can have more than one account_num, and an account_num can be associated with more than one account_id. See additional details in the *Accounts File* description. Variable type: Character; Variable Length: 9; Number of observations with missing values: 0; Number of unique non-missing values: 57,407.

4. **third_party_allowed_flag** - An indicator that shows whether the job posting accepts applications from 3rd parties. This may change over the course of a job posting's lifetime. Variable type: Character; Variable Length: 12; Number of observations with missing values: 0; Number of unique non-missing values in the Jobs File: 3.

   o No 3rd Party: Posting does not accept applications from third parties.
   o 3rd Party OK: Posting does accept applications from third parties.
   o Unknown: The client has not specified a preference regarding applications from 3rd parties.

Vacancy Days by Company Type and Third Party Allowed Flag							
Company Type							
Third Party Allowed Flag	Recruitment Firm		Direct Hire		Other		Total
	Vacancy Days	%	Vacancy Days	%	Vacancy Days	%	%
3rd Party OK	13,623,098	7.6	32,977,285	18.4	93,550	0.1	26.0
No 3rd Party	54,705,635	30.5	68,284,116	38.0	2,693,234	1.5	70.0
Unknown	2,839,845	1.6	4,247,414	2.4	110,000	0.1	4.0
Total	71,168,578	39.6	105,508,815	58.8	2,896,784	1.6	100.0

| Vacancy Days Share by Contract Terms and Third Party Allowed Flags			
	Corp-to-Corp Contract (Tax Terms)		
Third Party Allowed Flag	No	Yes	Missing
3rd Party OK	0.4	25.3	0.4
No 3rd Party	67.2	0.3	2.5
Unknown	0.0	0.0	4.0
Total	67.6	25.6	6.8

Note: Postings are categorized as "Corp-to-Corp" if the posting's tax terms, in the Jobs File, mention Con_Hire_Corp or Con_Corp contracts. Postings with any other tax terms are grouped as "Not Corp-to-Corp". Tax terms are fixed across a postings' vacancy days. Clients can change a posting's third party allowed flag on different vacancy days.

5. **date** – Date for the vacancy day in a year-month-day (e.g. 2017-05-22) format. Variable type: Date; Number of observations with missing values: 0; Number of unique non-missing values: 2,192.

6. **week** – Date of the first day of the week for each active vacancy day in a year-month-day format. Weeks start on Mondays. For example, all vacancy days within May 22, 2017 and May 28, 2017 will have a value of 2017-05-22 for the week variable. This variable allows grouping postings' activity on a weekly basis. Variable type: Date; Number of observations with missing values: 0; Number of unique non-missing values: 314.

7. **month** - Date of the first day of the month in a year-month-day format. For example, all vacancy days on May 2017 will have a value of 2017-05-01 in the month variable. This variable allows grouping postings' activity on a monthly basis. Variable type: Date; Number of observations with missing values: 0; Number of unique non-missing values: 72.

8. **days_since_first_active_date** – Number of days since the job (job_id) was first posted. It is the difference between the vacancy day date (day of the observation) and first_active_date. This variable does not discount inactive days but rather just indicates the total number of calendar days (active and inactive) elapsed since the date when the posting was first online. Variable type: Numeric; Number of observations with missing values: 0; Number of unique non-missing values: 2,192.
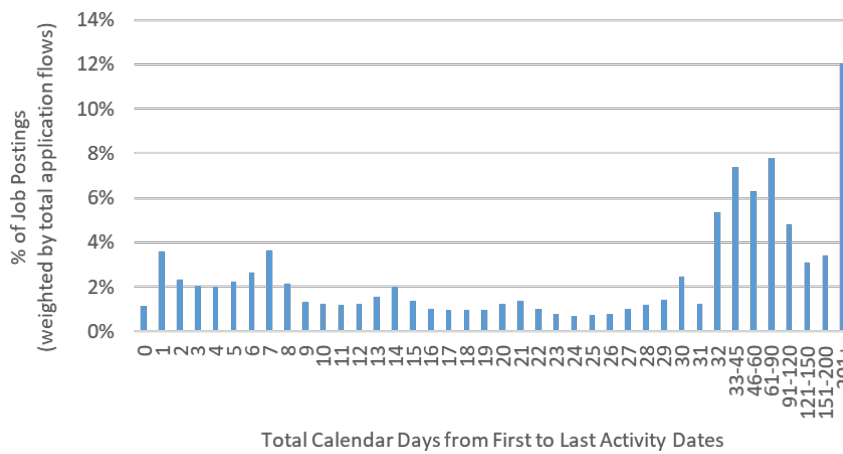
Job Postings by Maximum Number of Days Since First Active Date
Panel A: Equal weights for all job postings
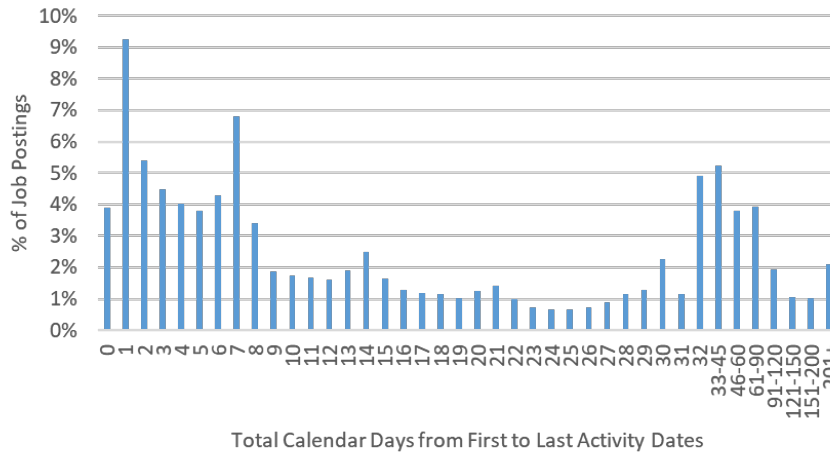
Panel B: Weighted by net posting duration (sum of active days)
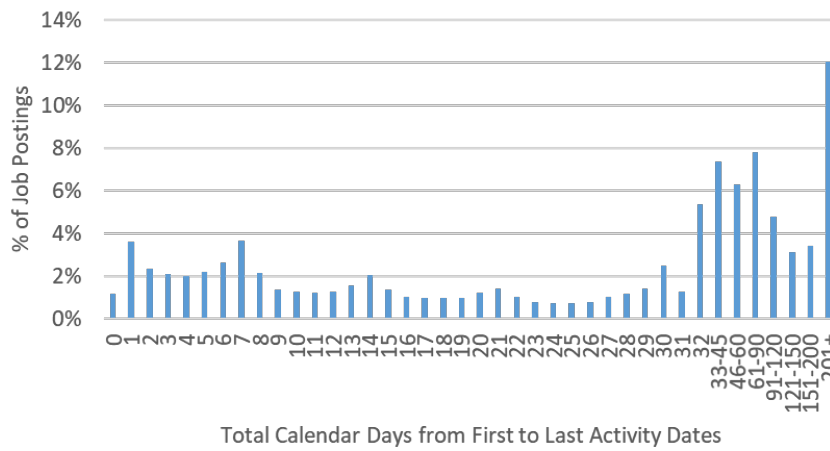
Panel C: Weighted by total applications received

Notes to Panels A, B, and C above:

1. Each job posting is associated with a unique job_id in the DHI data.
2. The graphs show the distribution of job postings (job_id's) by total number of days elapsed between the most recent date seen online (max(last_active_date)) and the first time the job was posted (first_active_date).
3. "Total calendar days from first to last activity dates" is equal to zero for job postings that are removed on the same day as the first active date, and never reposted.
4. (Panel B only) Active days online equal the number of calendar days when the posting is visible to potential applicants (sum of vacancy days). For postings that do not have any offline spells, net posting duration is equal to the total calendar days from first to last activity dates plus 1. For postings with offline spells, it is equal to the total calendar days from first to last activity dates minus the number of days, between first and last active dates, when the posting is not available for candidates to view plus 1.
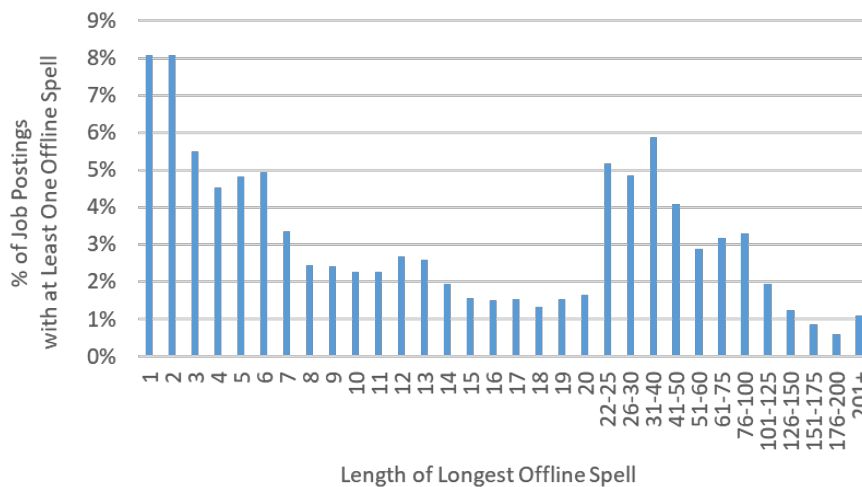
9. **days_since_last_seen_online** – Number of days since the posting was last active. For consecutive posting days, this variable is equal to 1. For job postings with any offline spells, on the date of reposting, this variable shows the length of the offline spell. days_since_last_seen_online is missing on the first active date. 89% of all postings do not have any offline spells. 3.9% of postings are taken offline on the same day they were first posted, and then are never re-posted. Variable type: Numeric. Number of observations with missing values: 7,838,006 (one for each new job posting). Number of unique non-missing values: 359.

Longest Offline Spell Distribution (Postings with at least one offline spell)	
Minimum	2
1st Pctl	2
10th Pctl	3
25th Pctl	5
50th Pctl	13
75th Pctl	32
90th Pctl	71
99th Pctl	206
Maximum	362
Mean	27.82

Std Dev	40.20
N	899,699

Notes: The table shows percentiles for each job posting longest offline spell. An offline spell is defined as the length of time (measured in calendar days) between non-consecutive vacancy days. It excludes postings that did not experience offline spells.

Job Posting Distribution by Length of Longest Offline Spell
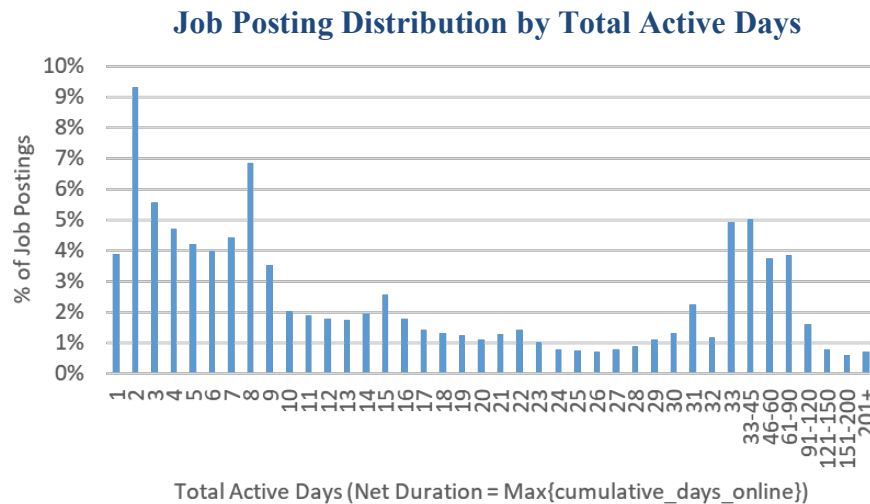**(Postings with at least one offline spell)**

Note: The length of the longest offline spell equals the maximum value for days_since_last_seen_online across all vacancy days for each posting minus one.

10. **cumulative_days_online** – Cumulative sum of the active calendar days, net of offline spells, for each job posting. This variable takes a value of 1 on the posting's first active date. Variable Type: Numeric. Number of observations with missing values: 0; Number of unique non-missing values: 366.

Total Active Days Distribution	
(Max. Value of Cumulative Days Online)	
Minimum	1
1st Pctl	1
10th Pctl	2
25th Pctl	5
50th Pctl	11
75th Pctl	30

90th Pctl	49
99th Pctl	173
Maximum	2,192
Mean	22.91
Std Dev	39.23
N	7,838,006

Job Posting Distribution by Total Active Days

Total Active Days (Net Duration = Max{cumulative_days_online})

Note: A posting is active if it is online and visible to applicants at any point during the day. Total active days (or net posting duration) is the sum of calendar days where the posting was active.

11. **time_online_in_seconds** – Number of seconds the job_id was active (i.e., visible to potential applicants) during each of its days online. With the exception of dates on which the job is posted or taken offline, this value is usually 86,400 seconds, that is, job postings are visible to applicants during the whole day (24 hours) on most vacancy days. There are 271 vacancy days (134 distinct job_id's) that, potentially incorrectly, record 0 seconds online. 13 of these vacancy days have positive application flows on these "zero seconds" days. 65 of them have positive job views. Variable type: Numeric. Number of observations with missing values: 0; Number of unique non-missing values: 86,401.

Statistics for Time Online in Seconds per Vacancy Day	
Minimum	0
1st Pctl	16,206
10th Pctl	66,552
25th Pctl	86,400
50th Pctl	86,400
75th Pctl	86,400
90th Pctl	86,400
99th Pctl	86,400
Maximum	86,400
Mean	81,368.99
Std Dev	15,581.03
N	179,574,177

12. **job_views** – Number of times the posting was viewed by jobseekers on each active date. This variable reports the number of views during the day, not the number of unique viewers. Note: Starting in April 2017, the job_views variable is missing. Variable type: Numeric. Number of observations with missing values: 20,861,080; Number of unique non-missing values: 3,456

Statistics for Job Views per Vacancy Day (Jan. 2012 – Mar. 2017)	
Minimum	0
1st Pctl	0
10th Pctl	3
25th Pctl	5
50th Pctl	9
75th Pctl	16
90th Pctl	26
99th Pctl	79
Maximum	9,962
Mean	13.49
Std Dev	24.36
N	158,713,097

Job Posting Distribution by Total Views Received Over Posting Duration
(Jobs First Posted Before March 1st, 2017)
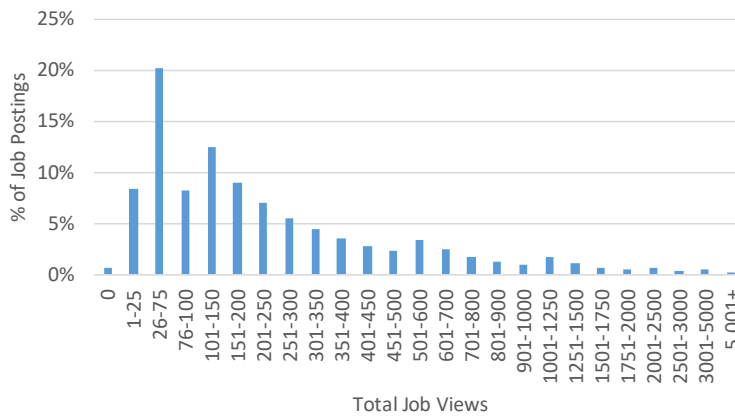
Notes: Total views is the sum of views received over all vacancy days for each posting. The sample is restricted to jobs first posted between January 1, 2012 and February 28, 2017.

Job Views Distribution by Number of Calendar Days
Since Job Was First Posted (Jan. 2012 – Mar. 2017)
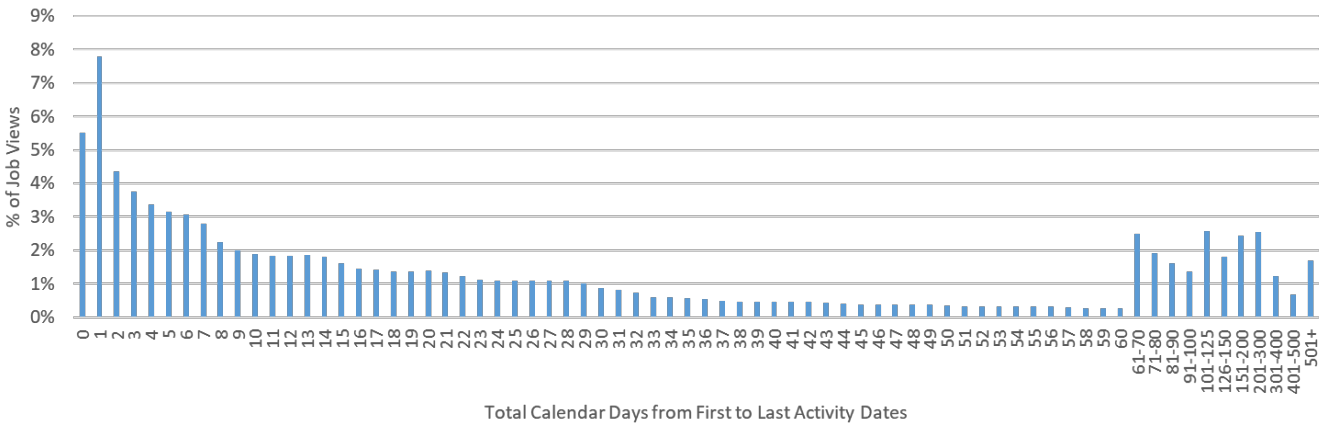
Notes: 0 on the horizontal axis refers to views on the first active calendar day. The sample includes all views for vacancy days between January 1, 2012 and March 31, 2017.

13. **email_applies –** Number of completed applications via email per vacancy day. This variable is equal to zero if the client chose to accept URL applications on that vacancy day or if the client chose email applications but the posting did not receive any applicants that day. Most postings only allow for one type of applications throughout their posting duration. Variable type: Numeric. Number of observations with missing values: 0; Number of unique non-missing values: 361.

Email Applications per Vacancy Day	
Minimum	0
1st Pctl	0
10th Pctl	0
25th Pctl	0
50th Pctl	0
75th Pctl	1
90th Pctl	2
99th Pctl	11
Maximum	649
Mean	0.81
Std Dev	3.66
N	120,331,313

Notes: This table excludes all applications flows for job postings that received more URL than email applications.

14. **email_applies_from_us** - Number of completed applications via email by job seekers with IP addresses in the United States during the day. Variable type: Numeric. Number of observations with missing values: 0; Number of unique non-missing values: 116.

The following statistics refer to email applications per vacancy day, from applicants within the US, for postings that receive applications via email.

Email Applications from Applicants in the US per Vacancy Day	
Minimum	0
1st Pctl	0
10th Pctl	0
25th Pctl	0
50th Pctl	0
75th Pctl	0
90th Pctl	1
99th Pctl	5
Maximum	299

Mean	0.42
Std Dev	1.21
N	120,331,313

Notes: This table excludes all applications flows for job postings that received more URL than email applications.

15. **email_applies_from_outside_us -** Number of completed applications via email by job seekers with IP addresses outside the United States for each active vacancy day. Variable type: Numeric. Number of observations with missing values: 0; Number of unique non-missing values: 309.

Email Applications from Applicants Outside the US per Vacancy Day	
Minimum	0
1st Pctl	0
10th Pctl	0
25th Pctl	0
50th Pctl	0
75th Pctl	0
90th Pctl	0
99th Pctl	7
Maximum	648
Mean	0.36
Std Dev	2.83
N	120,331,313

Notes: This table excludes all applications flows for job postings that received more URL than email applications.

16. **email_applies_from_1st_party -** Number of completed applications via email per vacancy day by job seekers submitting the application on their own behalf (in contrast to 3rd party applications who apply on behalf of someone else). Variable type: Numeric. Number of observations with missing values: 0; Number of unique non-missing values: 94.

Email 1st Party Applications per Vacancy Day	
Minimum	0
1st Pctl	0
10th Pctl	0
25th Pctl	0
50th Pctl	0
75th Pctl	0
90th Pctl	1
99th Pctl	4
Maximum	329
Mean	0.30
Std Dev	0.90
N	120,331,313

Notes: This table excludes all applications flows for job postings that received more URL than email applications.

17. **url_applies:** Number of viewers redirected to an external application URL. Variable type: Numeric. Number of observations with missing values: 0; Number of unique non-missing values: 303.

URL Applications per Vacancy Day	
Minimum	0
1st Pctl	0
10th Pctl	0
25th Pctl	0
50th Pctl	0
75th Pctl	0
90th Pctl	1
99th Pctl	6
Maximum	895
Mean	0.52
Std Dev	2.14

N	59,242,864

19. **url_applies_from_us -** Number of viewers with IP addresses within the United States redirected to an external application URL. Variable type: Numeric. Number of observations with missing values: 0; Number of unique non-missing values: 150.

URL Applications from Applicants in the US per Vacancy Day	
Minimum	0
1st Pctl	0
10th Pctl	0
25th Pctl	0
50th Pctl	0
75th Pctl	0
90th Pctl	1
99th Pctl	5
Maximum	892
Mean	0.38
Std Dev	1.12
N	59,242,864

20. **url_applies_from_outside_us -** Number of viewers with IP addresses from outside the United States redirected to an external application URL. Variable type: Numeric. Number of observations with missing values: 0; Number of unique non-missing values: 241.

URL Applications from Applicants Outside the US per Vacancy Day	
Minimum	0
1st Pctl	0
10th Pctl	0

25th Pctl	0
50th Pctl	0
75th Pctl	0
90th Pctl	0
99th Pctl	3
Maximum	376
Mean	0.14
Std Dev	1.47
N	59,242,864

Notes: This table includes applications flows for job postings that received more URL than email applications.

21. **url_applies_from_1st_party -** Number of viewers, applying on their own behalf, redirected to an external application URL per vacancy day. Variable type: Numeric. Number of observations with missing values: 0; Number of unique non-missing values: 150.

URL 1st Party Applications per Vacancy Day	
Minimum	0
1st Pctl	0
10th Pctl	0
25th Pctl	0
50th Pctl	0
75th Pctl	0
90th Pctl	1
99th Pctl	5
Maximum	892
Mean	0.38
Std Dev	1.12
N	59,242,864

Notes: This table includes applications flows for job postings that received more URL than email applications.
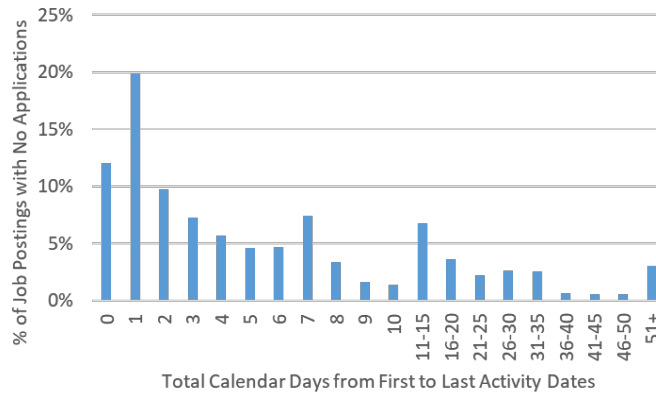
Other Relevant Statistics about Application Flows from the Activity File

| Share of Job Postings with Zero Applicants by Year of First Active Date					
2012	**2013**	**2014**	**2015**	**2016**	**2017**
20%	21%	19%	16%	17%	19%

Notes: This table considers only job postings included in the Activity File.

**Share of Job Postings with Zero Applications by Total Active Days**

Notes: The graph shown above includes only postings that did not receive any applications on any of the days it was active.

Job Postings Distribution by Total Applications Received
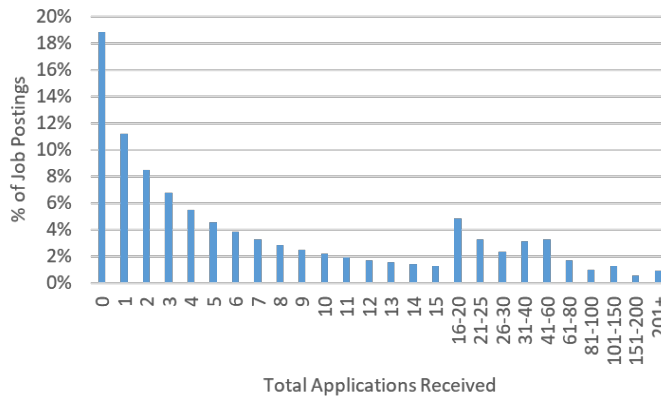
Percent of Applications Received by Job Posting Age

Calendar Days Elapsed from First Activity Date to Application
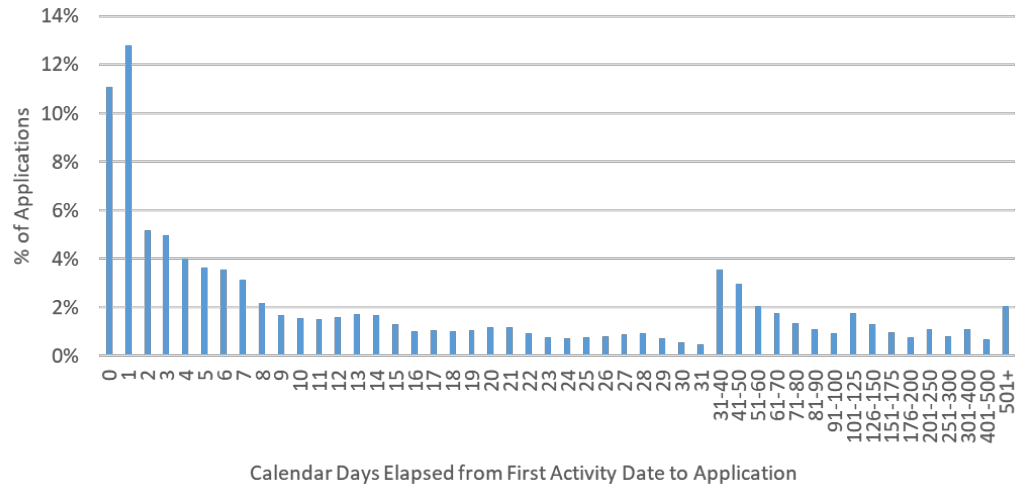
The *Detailed Applications* File

File Name: detailed_applications.sas7bdat
File Size: 45.1 GB
Encoding: latin1 Western (ISO)
Total Number of Observations: 128,952,237
Total Number of Variables: 12

The *Detailed Applications* File includes information about each application received by each job posting (described in the *Jobs* File) posted by DHI clients (whose information is contained in the *Accounts* File). For each application-job posting pair, the *Detailed Applications* File describes the exact date-time stamp when the application occurred, the type of application (email vs. URL), the applicant's job title, location (US or non-US), and work authorization status. This dataset also indicates whether the application is sent directly by the applicant (first_party_applicant_flag) or not, and whether the job allows third party applications.

Variable Order, Definitions, Admissible Values, and Summary Statistics:

1. **application_date_dt** – Date the application was received in a year-month-day (e.g. 2017-05-22) format. Variable type: Date; Number of observations with missing values: 0; Number of unique non-missing values: 2,189.

Application Distribution by Year

Year	No. of Applications	% of Applications
2012	12,656,143	9.8%
2013	13,435,151	10.4%
2014	13,095,785	10.2%
2015	23,857,791	18.5%
2016	34,337,496	26.6%
2017	31,569,871	24.5%

2. **application_date_time** - Date and time the application was received in a year-month-day hh:mm:ss (e.g. 2017-05-22 14:34:51) format. Variable type: Date; Number of observations with missing values: 0; Number of unique non-missing values: 67,325,379.

3. **job_id -** A unique alpha-numeric identifier for each job posting. This variable can be used to merge the *Jobs, Activity,* and *Detailed* Files. Variable type: Character; Variable Length: 32; Number of observations with missing: 0; Number of unique non-missing values: 6,366,373.

4. **application_type** – Indicates whether the application was received through email (coded as "MAIL" in the data) or if the applicant was redirected to an external URL. Variable type: Character; Variable Length: 4; Number of observations with missing: 0; Number of unique non-missing values: 2.

Application Distribution by Channel		
Channel	**No. of applications**	**% of applications**
Email	97,641,357	75.7
URL	31,310,880	24.3

5. **third_party_allowed_flag** – Indicates whether the job posting accepts applications from 3rd parties. This may change over the course of a job posting's duration. Variable type: Character; Variable Length: 12; Number of observations with missing values: 0; Number of unique non-missing values: 3.

6. **applicant_third_party_flag** – Indicates whether the applicant is submitting the application on their own behalf (1st party application) or if someone else is submitting the application on behalf of the potential employee (3rd party application). 3.1% of all applications in the *Detailed* File are 3rd party applications submitted to postings that do not allow such applications. Variable type: Character; Variable Length: 9; Number of observations with missing values: 0; Number of unique non-missing values: 3.

Application Distribution by Third-Party Requirements				
Does the posting Allows 3rd Party Applications?	**Application Type**			
	1st Party	**3rd Party**	**Unknown**	**Total**
No 3rd Party	8,953,562 6.9%	3,999,493 3.1%	682,429 0.5%	13,635,484 10.6%
3rd Party OK	4,478,071 3.5%	29,488,901 22.9%	439,858 0.3%	34,406,830 26.7%
Unspecified	22,120,531 17.2%	40,942,552 31.8%	17,846,840 13.8%	80,909,923 62.7%
Total	35,552,164 27.6%	74,430,946 57.7%	18,969,127 14.7%	128,952,237 100.0%

Notes: The table above shows the distribution of applications by postings' third-party requirements and applicants' compliance with these requirements. Each column classifies applications based on the application type, that is, whether the application is submitted directly by the applicant (1st party), on behalf of someone else (3rd party), or if this information is not available (Unknown). Each row indicates whether the posting that received the application allows for 3rd party applications or not, or whether the posting does not specify constraints with regards to 3rd party applicants (Unspecified).

7. **work_authorization** – Applicants' work authorization status. An applicant's work authorization can vary through time. Therefore, different applications by the same applicant can have different values for this variable. However, most applicants (98%) have the same work authorization status across all their applications. Variable type: Character; Variable Length: 30; Number of observations with missing values: 0; Number of unique non-missing values: 9.

Application Distribution by Applicant Work Authorization Status		
Work Authorization	**No. of Applications**	**Share of Applications**
Have H-1 Visa	52,636,861	40.8
U.S. Citizen	28,416,941	22.0
Unknown	21,103,374	16.4
Employment Authorization	12,818,209	9.9
Green Card Holder	9,342,086	7.2
Need H-1 Visa	3,376,386	2.6
TN Permit Holder	821,333	0.6
Canadian Citizen	436,831	0.3
Authorized to work in the US	216	0.0

8. **applicant_position_title** - Applicant's self-reported job title. The applicant could refer to his current job title, qualifications, previous job title, or even desired position. Variable type: Character; Variable Length: 247; Number of observations with missing values: 22,453,830; Number of unique non-missing values: 999,275.

9. **applicant_id** - A unique alpha-numeric identifier for each applicant. This variable can be used to identify all applications submitted by any applicant. Variable type: Character; Variable Length: 12; Number of observations with missing: 0; Number of unique non-missing values: 5,670,297.

10. **application_id** - A unique alpha-numeric identifier for each application and, hence, for each observation in the Application Details file. Variable type: Character; Variable Length: 10;

Number of observations with missing: 0; Number of unique non-missing values: 128,952,237.

11. **applicant_country** – This variable indicates whether the application comes from an IP address within the United States (US) or not (Non-US). 86% of all applicants have IP addresses within the United States, and they submit 57% of all applications Variable type: Character; Variable Length: 6; Number of observations with missing: 0; Number of unique non-missing values: 2.

12. **seconds_since_previous_apply** – Number of second elapsed since the applicant's previous application. This variable is missing for each applicant's first application. Variable type: Numeric; Number of observations with missing: 5,735,515; Number of unique non-missing values: 100,001.