

Worksheet - Handling form data

Overview

Survey on workplace happiness: 630 diverse participants, multiple industries, collected info. and happiness scores (0-10) in data centric roles (6 happiness metrics)

<https://www.kaggle.com/datasets/deepalisukhdeve/data-professional-survey/data>

Goal

Extract insights from mixed, average quality survey data (qualitative & quantitative)

Sample results

in Canada

Overall happiness scores: 4.4 (already in data), 5.5 (transitioned to data)

Highest happiness gap is in 2 main metrics: work/life balance, continuous learning (higher in those who transitioned to data by 2+ points)

Average salaries: 99 kCAD (already in data), 88 kCAD (transitioned to data)

(see dashboard sheets for visualisations)

Data cleaning steps

Columns

Delete empty columns (country where survey completed, etc)

Rename columns (shorter names)

Reorder columns (more coherent themes)

Format times & dates (yyyy-mm-dd)

Country

Correct countries (typos, full name, proper case)

Ethnicity

Group all Other as one (many unique entries)

Title

For Other (Specify) responses:

Tally keywords (BI, Manager, etc.)

Create popular categories (top 3, 10+ members)

Rename unpopular categories as "Other"

Industry

For Other (Specify) responses:

Tally keywords (Automotive, Retail, Fashion, etc.)

Create popular categories (top 7, 10+ members)

Rename unpopular categories as "Other"

Average salary

Extract min and max from range given

Compute average

The 225k+ tier gets an average of 225k + average(max-min) from other tiers

Switched to data

Convert text (Yes/No) to bool

Switch difficulty

Convert text to integer scale (1-5)

Favourite programming language

For Other (Specify) responses:

Tally keywords (SQL, SAS, C#, learning, etc.)

Create popular categories (top 3, 10+ members)

Rename unpopular categories as "Other"

Important work metric

For Other (Specify) responses:

Tally keywords (learn, growth, decision, passion, etc.)

Create popular categories (top 3, 10+ members)

Rename unpopular categories as "Other"

Remembered to TRIM() all text entries

Rows

Deleted

Any missing scores (18 rows)

Outliers

Computed z scores for numeric columns (see note)

Set z score threshold at 3

Assessed outlier subset:

Deleted

1h+ survey durations (2 rows)

Included

Salary outliers (valid values)

Age outliers (valid survey takers)

Note**Outlier detection**

The handmade formula below detects a table's numeric outliers.

```
=(ABS((((FILTER(T,ISNUMBER(INDEX(T,1,SEQUENCE(,COLUMNS(T)))))-(TRANSPOSE(MMULT(TRANSPOSE(FILTER(T,ISNUMBER(INDEX(T,1,SEQUENCE(,COLUMNS(T))))),SEQUENCE(ROWS(FILTER(T,ISNUMBER(INDEX(T,1,SEQUENCE(,COLUMNS(T))))),1/ROWS(FILTER(T,ISNUMBER(INDEX(T,1,SEQUENCE(,COLUMNS(T))))),0)))))/SQRT(TRANSPOSE(MMULT(TRANSPOSE((FILTER(T,ISNUMBER(INDEX(T,1,SEQUENCE(,COLUMNS(T)))))-(TRANSPOSE(MMULT(TRANSPOSE(FILTER(T,ISNUMBER(INDEX(T,1,SEQUENCE(,COLUMNS(T))))),SEQUENCE(ROWS(FILTER(T,ISNUMBER(INDEX(T,1,SEQUENCE(,COLUMNS(T))))),1/ROWS(FILTER(T,ISNUMBER(INDEX(T,1,SEQUENCE(,COLUMNS(T))))),0))))^2),SEQUENCE(ROWS((FILTER(T,ISNUMBER(INDEX(T,1,SEQUENCE(,COLUMNS(T))))-(TRANSPOSE(MMULT(TRANSPOSE(FILTER(T,ISNUMBER(INDEX(T,1,SEQUENCE(,COLUMNS(T))))),SEQUENCE(ROWS(FILTER(T,ISNUMBER(INDEX(T,1,SEQUENCE(,COLUMNS(T))))),1/ROWS(FILTER(T,ISNUMBER(INDEX(T,1,SEQUENCE(,COLUMNS(T))))),0))))^2,,1/ROWS(FILTER(T,ISNUMBER(INDEX(T,1,SEQUENCE(,COLUMNS(T))))),0))))))>3)
```

> For a table named T, first detect numeric columns, then return TRUE or FALSE matching outlier entries (based on a z-score cutoff of 3)

Since the release of GenAI, this formula was iteratively improved (legibility and header handling), mainly using Gemini, to:

```
=ARRAYFORMULA(
  LET(
    effective_table_name, IF(called_with_headers, OFFSET(table_name, 1, 0, ROWS(table_name)-1, COLUMNS(table_name)), table_name),
    is_numeric_column, ISNUMBER(INDEX(effective_table_name, 2, 0)),
    filtered_data, IFERROR(FILTER(effective_table_name, is_numeric_column) * 1, 0),
    num_rows_filtered, ROWS(filtered_data),
    num_cols_filtered, COLUMNS(filtered_data),
    display_headers,
    IF(called_with_headers,
      FILTER(INDEX(table_name, 1, 0), is_numeric_column),
      IF(num_cols_filtered > 0, "numeric_column_" & SEQUENCE(1, num_cols_filtered, 1, 1), "")
    ),
    mean_data, TRANSPOSE(MMULT(TRANSPOSE(filtered_data), SEQUENCE(num_rows_filtered, 1, 1/num_rows_filtered, 0))),
    diff_from_mean, filtered_data - mean_data,
    std_dev_data, SQRT(TRANSPOSE(MMULT(TRANSPOSE(diff_from_mean^2), SEQUENCE(num_rows_filtered, 1, 1/num_rows_filtered, 0)))),
    abs_z_score, ABS(diff_from_mean / std_dev_data),
    {display_headers; abs_z_score > z_score}
  )
)
```

The result is a named function with the following syntax:

LIST_OUTLIERS(table_name, z_score, called_with_headers)

> An example call (in a single cell): =LIST_OUTLIERS(FullTable, 1.9, True)

> Its output is included in the last bonus sheet

Possible improvements:

More thorough verification that no outliers are missed

Time to complete the analysis

On and off throughout two weeks (~5 half-days)

Google Sheets link

<https://docs.google.com/spreadsheets/d/1Vbmel8KICPcz93UZbATaafTOSsaVNQnmlcwCa9dfnHQ/edit?usp=sharing>

Unique ID	Date Taken (Am	Time Taken (Am	Time Spent	Q1 - Which Title	Q2 - Did you swi	Q3 - Current Year	Q4 - What Indus	Q5 - Favorite Pri	Q6 - How Happy	Q6 - How Happy	Q6 - How Happy	Q6 - How Happy	Q6 - How Happy	Q7 - How difficu	Q8 - If you were	Q9 - Male/Fema	Q10 - Current	Aj Q11 - Which	Coi Q12 - Highest	Le Q13 - Ethnicity
62a33b3db4da2	6/10/2022	08:38	00:00:44	Data Analyst	Yes	106k-125k	Healthcare	Python	9	9	7	5	2	1	7 Very Difficult	Remote Work	Male	26	United States	White or Caucasian
62a33ba1bae91c	6/10/2022	08:40	00:01:30	Data Analyst	No	41k-65k	Finance	R	1	2	5	5	1	3 Very Difficult	Remote Work	Male	36	Canada	Asian or Asian American	
62a33c2cbcb6861	6/10/2022	08:42	00:02:18	Data Engineer	No	0-40k	Other (Please Sp	Python	0	8	7	5	7	7 Easy	Good Work/Life	Male	23	Other (Please Specify):Nigeria	Black or African American	
62a33c8624a26c	6/10/2022	08:43	00:02:10	Other (Please Sp	Yes	150k-225k	Finance	R	10	6	7	10	7	10 Very Easy	Remote Work	Female	35	Canada	White or Caucasian	
62a33c91f3072d	6/10/2022	08:44	00:01:51	Data Analyst	Yes	41k-65k	Healthcare	R	1	4	4	4	0	1 Difficult	Better Salary	Male	44	United States	Black or African American	
62a33cb6cf2555	6/10/2022	08:44	00:02:34	Data Analyst	Yes	0-40k	Other (Please Sp	Python	2	3	3	3	3	2 Difficult	Good Work/Life	Male	33	Other (Please Specify):Republic d	Black or African American	
62a33cb72e54c4	6/10/2022	08:44	00:01:15	Data Scientist	Yes	0-40k	Finance	Python	0	6	6	6	6	5 Difficult	Better Salary	Male	24	Other (Please Specify)	Black or African American	
62a33cd30f8c85	6/10/2022	08:45	00:01:25	Data Engineer	Yes	125k-150k	Other (Please Sp	Other:SQL	10	5	5	5	5	9 Neither easy nor	Remote Work	Male	35	United States	Asian or Asian American	
62a33cd3cf2555	6/10/2022	08:45	00:02:10	Data Analyst	Yes	86k-105k	Healthcare	R	4	4	4	4	4	5 Neither easy nor	Better Salary	Male	35	United States	Asian or Asian American	
62a33cd8bcb6861	6/10/2022	08:45	00:01:27	Data Analyst	Yes	41k-65k	Telecommunicat	Python	2	0	3	3	0	4 Very Difficult	Better Salary	Male	28	United States	Hispanic or Latino	
62a33ce9f18134c	6/10/2022	08:45	00:01:29	Data Analyst	Yes	66k-85k	Other (Please Sp	Python	10	10	8	8	8	10 Neither easy nor	Remote Work	Male	37	Canada	White or Caucasian	
62a33cf30a77c1	6/10/2022	08:45	00:02:31	Data Analyst	Yes	0-40k	Other (Please Sp	Python	3	8	10	10	7	8 Neither easy nor	Better Salary	Male	32	Other (Please Specify):Serbia	White or Caucasian	
62a33d15f408ba	6/10/2022	08:46	00:03:20	Data Analyst	Yes	0-40k	Tech	R	1	10	9	9	5	10 Neither easy nor	Better Salary	Male	34	Other (Please Specify):Argentina	Other (Please Specify):Latino with Italian roots	
62a33d1ebae91c	6/10/2022	08:46	00:00:55	Data Scientist	Yes	0-40k	Education	Python	3	10	10	6	7	8 Easy	Good Work/Life	Male	30	United Kingdom	White or Caucasian	
62a33d4624a26c	6/10/2022	08:47	00:01:24	Data Analyst	No	41k-65k	Construction	R	4	5	5	5	4	5 Neither easy nor	Good Culture	Male	31	United States	White or Caucasian	
62a33d5c0f8c85	6/10/2022	08:47	00:00:47	Data Analyst	Yes	41k-65k	Finance	Python	4	10	7	7	7	5 Difficult	Good Culture	Male	29	United States	White or Caucasian	
62a33d850f8c85	6/10/2022	08:48	00:01:06	Data Analyst	Yes	0-40k	Tech	Python	1	2	4	2	2	2 Neither easy nor	Better Salary	Male	22	Other (Please Specify):Niger	Black or African American	
62a33d9bfb8f756	6/10/2022	08:48	00:01:04	Student/Looking	Yes	0-40k	Other (Please Sp	Python	1	3	2	2	2	1 Difficult	Better Salary	Male	23	United States	Black or African American	
62a33dbdd8c00c	6/10/2022	08:49	00:01:05	Student/Looking	No	41k-65k	Tech	R	7	8	8	6	8	3 Very Difficult	Better Salary	Female	26	Canada	Asian or Asian American	
62a33de124a26c	6/10/2022	08:49	00:01:21	Data Analyst	No	0-40k	Other (Please Sp	Python	1	1	1	1	2	0 Very Difficult	Better Salary	Male	28	Other (Please Specify):Nigeria	Black or African American	
62a33dedd8c00c	6/10/2022	08:49	00:01:23	Data Analyst	Yes	41k-65k	Finance	Python	5	5	5	5	5	5 Easy	Remote Work	Male	29	Other (Please Specify):Kenya	Black or African American	
62a33dfbdf8c00c	6/10/2022	08:50	00:01:35	Data Analyst	Yes	106k-125k	Tech	Other:Mostly us	10	5	10	6	10	10 Neither easy nor	Remote Work	Female	31	United States	White or Caucasian	
62a33e38f3072c	6/10/2022	08:51	00:01:25	Data Analyst	Yes	0-40k	Healthcare	Python	5	6	8	8	7	1 Neither easy nor	Remote Work	Male	28	Other (Please Specify):Azerbaijan	White or Caucasian	
62a33e54dd8c00c	6/10/2022	08:51	00:01:47	Data Analyst	No	0-40k	Finance	Python	4	4	4	4	4	4 Neither easy nor	Better Salary	Male	32	Other (Please Specify):Argentina	Hispanic or Latino	
62a33e62575fa3	6/10/2022	08:51	00:02:14	Data Analyst	No	0-40k	Tech	Python	7	8	8	8	3	8 Easy	Better Salary	Male	24	Other (Please Specify):Sudan	Black or African American	
62a33e63cec54f	6/10/2022	08:51	00:02:21	Data Analyst	No	0-40k	Tech	Python	4	6	6	6	7	7 Easy	Good Work/Life	Male	31	India	Asian or Asian American	
62a33e72bae91c	6/10/2022	08:52	00:02:07	Data Analyst	No	66k-85k	Other (Please Sp	Other:SQL	7	10	5	5	5	2 Difficult	Other (Please Sp	Male	23	United States	White or Caucasian	
62a33e890f8c85	6/10/2022	08:52	00:01:20	Data Analyst	Yes	106k-125k	Other (Please Sp	Python	1	6	9	1	0	2 Difficult	Remote Work	Female	33	United States	Asian or Asian American	
62a33e8bf8d5ef	6/10/2022	08:52	00:01:34	Data Analyst	No	86k-105k	Healthcare	Other:Qlik sense	6	7	7	7	7	4 Very Easy	Better Salary	Male	43	United States	White or Caucasian	
62a33e90cec54f	6/10/2022	08:52	00:01:34	Student/Looking	Yes	0-40k	Other (Please Sp	R	0	0	5	0	0	0 Very Difficult	Remote Work	Female	23	Other (Please Specify):Nigeria	Other (Please Specify):African	
62a33e9df8f756	6/10/2022	08:52	00:01:47	Student/Looking	Yes	41k-65k	Finance	Python	10	7	10	10	5	3 Very Difficult	Good Work/Life	Male	31	United States	Black or African American	
62a33eb6bcb6861	6/10/2022	08:53	00:01:38	Data Analyst	Yes	66k-85k	Finance	Python	5	6	7	5	4	4 Neither easy nor	Other (Please Sp	Male	31	United Kingdom	White or Caucasian	
62a33ee0bae91c	6/10/2022	08:53	00:03:30	Data Analyst	Yes	0-40k	Other (Please Sp	Python	8	7	8	7	7	7 Difficult	Good Work/Life	Male	25	India	Asian or Asian American	
62a33ee80f8c85	6/10/2022	08:54	00:01:37	Data Engineer	No	125k-150k	Tech	Other:SQL	8	10	10	9	9	9 Easy	Better Salary	Male	43	United States	White or Caucasian	
62a33ef6da77c1	6/10/2022	08:54	00:02:08	Data Analyst	No	41k-65k	Finance	Python	2	3	5	5	5	5 Difficult	Better Salary	Female	26	Canada	Other (Please Specify):South Asian	
62a33f0ef408ba	6/10/2022	08:54	00:03:54	Other (Please Sp	No	41k-65k	Other (Please Sp	Python	5	10	8	8	8	7 Neither easy nor	Remote Work	Male	32	Other (Please Specify):Japan	White or Caucasian	
62a33f1f8f756f	6/10/2022	08:54	00:02:28	Other (Please Sp	No	0-40k	Tech	Python	2	9	7	7	5	6 Difficult	Better Salary	Male	31	Other (Please Specify):Greece	White or Caucasian	
62a33f52bcb6861	6/10/2022	08:55	00:01:33	Data Analyst	Yes	86k-105k	Healthcare	Other:sql	3	5	6	4	4	5 Neither easy nor	Better Salary	Male	29	United States	White or Caucasian	
62a33f56cf2555	6/10/2022	08:55	00:00:56	Data Analyst	Yes	0-40k	Other (Please Sp	Python	3	1	3	2	3	1 Neither easy nor	Good Culture	Male	24	Other (Please Specify):Peru	Hispanic or Latino	
62a33f5a5d257a	6/10/2022	08:55	00:02:38	Student/Looking	Yes	41k-65k	Construction	Python	2	0	2	2	0	0 Very Difficult	Remote Work	Male	34	United States	White or Caucasian	
62a33f5cec549	6/10/2022	08:57	00:01:52	Data Analyst	Yes	41k-65k	Other (Please Sp	Other:Power bi	5	6	6	5	4	6 Very Difficult	Better Salary	Male	42	Other (Please Specify):Australia	Other (Please Specify):Indian	
62a33fcb575fa3c	6/10/2022	08:57	00:02:03	Student/Looking	Yes	0-40k	Other (Please Sp	JavaScript	0	0	0	0	0	0 Difficult	Remote Work	Female	40	United States	Asian or Asian American	
62a33fcd18134d	6/10/2022	08:57	00:02:32	Student/Looking	No	41k-65k	Telecommunicat	Python	3	1	3	4	0	3 Easy	Good Work/Life	Male	43	Other (Please Specify):Spain	White or Caucasian	
62a34006d8c00c	6/10/2022	08:58	00:02:39	Student/Looking	Yes	0-40k	Construction	Python	0	0	0	0	0	0 Very Difficult	Better Salary	Male	23	Other (Please Specify):Kenya	Other (Please Specify):Kenyan African	
62a34020575fa3	6/10/2022	08:59	00:01:11	Student/Looking	Yes	0-40k	Education	Python	5	5	3	5	2	6 Difficult	Better Salary	Male	23	Other (Please Specify):Japan	Asian or Asian American	
62a34038bd4a2f	6/10/2022	08:59	00:03:55	Data Analyst	Yes	86k-105k	Other (Please Sp	Python	8	6	2	3	6	7 Neither easy nor	Other (Please Sp	Male	28	United States	White or Caucasian	
62a3406b575fa3	6/10/2022	09:00	00:01:25	Data Analyst	No	66k-85k	Finance	Other:SQL	5	7	1	1	0	8 Neither easy nor	Better Salary	Male	32	United States	White or Caucasian	
62a340cab4da2f	6/10/2022	09:02	00:02:47	Data Analyst	Yes	66k-85k	Other (Please Sp	R	7	8	10	8	9	9 Difficult	Remote Work	Female	41	United States	White or Caucasian	
62a340d62e54c4	6/10/2022	09:02	00:00:56	Data Analyst	Yes	41k-65k	Finance	Python	5	6	6	6	6	6 Neither easy nor	Remote Work	Male	33	United Kingdom	High School	
62a340e10a77c1	6/10/2022	09:02	00:01:57	Other (Please Sp	Yes	150k-225k	Tech	Python	10	8	9	8	8	10 Easy	Good Work/Life	Male	38	United States	Hispanic or Latino	
62a340f5bcb6861	6/10/2022	09:02	00:02:32	Data Analyst	No	0-40k	Tech	Other:sql	8	8	8	8	5	10 Easy	Better Salary	Female	22	India	Other (Please Specify):Indian	
62a34117cec54f	6/10/2022	09:03	00:01:59	Student/Looking	Yes	0-40k	Other (Please Sp	R	1	3	6	6	6	9 Difficult	Remote Work	Male	36	Other (Please Sp	Bachelors	
62a3413a5d257f	6/10/2022	09:03	00:01:18	Data Engineer	No	66k-85k	Finance	Python	2	1	1	1	3	7 Easy	Remote Work	Female	28	United States	Masters	
62a341a1f3072c	6/10/2022	09:05	00:08:11	Student/Looking	Yes	0-40k	Other (Please Sp	Other:If SQL is categorised as a programming language then I will definitely say SQL. Since i am still le	3	7	7	7	7	7 Very Easy	Other (Please Sp	Male	30	Other (Please Specify):South Afric	Other (Please Specify):Black African, Zimbabwean Citizen	
62a341b0f8f756	6/10/2022	09:05	00:01:44	Data Analyst	Yes	66k-85k	Tech	Python	6	6	6	6	5	6 Difficult	Other (Please Sp	Female	24	United States	Bachelors	
62a341e8f8d5ef	6/10/2022	09:06	00:01:58	Other (Please Sp	Yes	41k-65k	Other (Please Sp	C/C++	3	6	7	7	0	3 Very Difficult	Better Salary	Male	29	United States	Bachelors	
62a34210cf2555	6/10/2022	09:07	00:02:15	Data Analyst	No	0-40k	Tech	Python	4	9	10	9	9	10 Easy	Better Salary	Male	29	Other (Please Sp	Associates	
62a3424eb4da2f	6/10/2022	09:08	00:02:04	Other (Please Sp	Yes	41k-65k	Other (Please Sp	Python	6	8	8	7	9	7 Neither easy nor	Good Work/Life	Male	38	Other (Please Sp	Masters	
62a34256bae91c	6/10/2022	09:08	00:01:46	Data Analyst	Yes	0-40k	Finance	Other: i don't knu	3	5	4	3	3	3 Difficult	Good Work/Life	Male	31	Other (Please Sp	Masters	
62a3426cf408ba	6/10/2022	09:09	00:01:33	Student/Looking	No	0-40k	Tech	Python	1	4	5	5	1	1 Difficult	Better Salary	Male	25	United States	Masters	
62a34272f8d5ef	6/10/2022	09:09	00:01:09	Data Analyst	No	66k-85k	Telecommunicat	Python	4	6	4	5	4	4 Neither easy nor	Better Salary	Male	24	United States	Masters	
62a34340bcb6861	6/10/2022	09:12	00:04:03	Other (Please Sp	Yes	41k-65k	Tech	Python	6	7	4	3	9	8 Difficult	Better Salary	Male	28	Other (Please Sp	Bachelors	
62a34383bd4da2f	6/10/2022	09:13	00:01:03	Student/Looking	No	0-40k	Tech	Python	7	5	8	0	0	0 Neither easy nor	Good Culture	Female	23	India	Bachelors	
62a34389cec54f	6/10/2022	09:13	00:01:01	Data Engineer	No	0-40k	Tech	R	2	3	2	5	5	7 Neither easy nor	Remote Work	Male	21	Other (Please Sp	Masters	
62a343a22e54c4	6/10/2022	09:14	00:02:43	Data Analyst	No	0-40k	Other (Please Sp	Python	4	4	4	4	5	6 Neither easy nor	Remote Work	Male	29	Other (Please Sp	Bachelors	
62a343d9f3072c	6/10/2022	09:15	00:03:15	Data Analyst	Yes	41k-65k	Tech	Python	5	6	7	7	7	6 Difficult	Better Salary	Male	25	United States	Bachelors	
62a34400d8c00c	6/10/2022	09:15	00:01:24	Data Analyst	No	41k-65k	Tech	Python	6	6	6	6	3	5 Difficult	Good Culture	Male	27	Other (Please Sp	Masters	
62a34425f8d5ef	6/10/2022	09:16	00:01:28	Data Analyst	Yes	41k-65k	Tech	Python	3	9	10	8	9	9 Easy	Better Salary	Male	23	India	Bachelors	
62a34449575fa3	6/10/2022	09:16	00:02:36	Other (Please Sp	No	86k-105k	Other (Please Sp	R	4	3	3	5	5	5 Difficult	Better Salary	Male	32	United States	Masters	
62a3445dcf2555	6/10/2022	09:17	00:02:41	Student/Looking	Yes	106k-125k	Healthcare	Python	5											

ID	survey_Date	survey_Time	survey_Duration	country	ethnicity	highest_Educati	gender	age	industry_from	title	avg_salary_KUSI	switched_To_De	switch_Difficult	favorite_Prograi	important_Word	happiness_Salar	happiness_Balai	happiness_Cow	happiness_Man	happiness_Upw	happiness_Cont	happiness_Avo	
62a33b3db4da2	2022-10-06	08:38:00	00:00:44	United States	White or Caucas	Unspecified	Male		26	Healthcare	Data Analyst	115.5	TRUE	5	Python	Remote Work	9	9	7	5	5	7	7.0
62a33ba1bae91	2022-10-06	08:40:00	00:01:30	Canada	Asian or Asian A	Unspecified	Male		36	Finance	Data Analyst	53	FALSE	5	R	Remote Work	1	2	5	2	1	3	2.3
62a33c2cbc861	2022-10-06	08:42:00	00:02:18	Nigeria	Black or African	Unspecified	Male		23	Energy	Data Engineer	20	FALSE	2	Python	Good Work/Life	0	8	7	5	7	7	5.7
62a33c8624a26	2022-10-06	08:43:00	00:02:10	Canada	White or Caucas	Unspecified	Female		35	Finance	Other	187.5	TRUE	1	R	Remote Work	10	6	7	10	7	10	8.3
62a33c91f3072d	2022-10-06	08:44:00	00:01:51	United States	Black or African	Unspecified	Male		44	Healthcare	Data Analyst	53	TRUE	4	R	Better Salary	1	4	4	4	0	1	2.3
62a33cb6cf2555	2022-10-06	08:44:00	00:02:34	Republic Democ	Black or African	Unspecified	Male		33	Other	Data Analyst	20	TRUE	4	Python	Good Work/Life	2	3	3	3	3	2	2.7
62a33cb72e54cf	2022-10-06	08:44:00	00:01:15	Other	Black or African	Unspecified	Male		24	Finance	Data Scientist	20	TRUE	4	Python	Better Salary	0	6	6	6	6	5	4.8
62a33cd30f8c85	2022-10-06	08:45:00	00:01:25	United States	Asian or Asian A	Unspecified	Male		35	Retail	Data Engineer	137.5	TRUE	3	SQL	Remote Work	10	5	5	5	5	9	6.5
62a33cd3cf2555	2022-10-06	08:45:00	00:02:10	United States	Asian or Asian A	Unspecified	Male		35	Healthcare	Data Analyst	95.5	TRUE	3	R	Better Salary	4	4	4	4	5	5	4.3
62a33cd8bc686f	2022-10-06	08:45:00	00:01:27	United States	Hispanic or Latin	Unspecified	Male		28	Telecommunicat	Data Analyst	53	TRUE	5	Python	Better Salary	2	0	3	5	0	4	2.0
62a33ce918134d	2022-10-06	08:45:00	00:01:29	Canada	White or Caucas	Unspecified	Male		37	Logistics	Data Analyst	75.5	TRUE	3	Python	Remote Work	10	10	8	8	8	10	9.0
62a33cf30a77c1	2022-10-06	08:45:00	00:02:31	Serbia	White or Caucas	Unspecified	Male		32	Other	Data Analyst	20	TRUE	3	Python	Better Salary	3	8	10	10	7	8	7.7
62a33d15f408bc	2022-10-06	08:46:00	00:03:20	Argentina	Other	Unspecified	Male		34	Tech	Data Analyst	20	TRUE	3	R	Better Salary	1	10	9	9	5	10	7.3
62a33d1ebae91	2022-10-06	08:46:00	00:00:55	United Kingdom	White or Caucas	Unspecified	Male		30	Education	Data Scientist	20	TRUE	2	Python	Good Work/Life	3	10	10	6	7	8	7.3
62a33d4624a26	2022-10-06	08:47:00	00:01:24	United States	White or Caucas	Unspecified	Male		31	Construction	Data Analyst	53	FALSE	3	R	Good Culture	4	5	5	5	4	5	4.7
62a33dc5cf8c85	2022-10-06	08:47:00	00:00:47	United States	White or Caucas	Unspecified	Male		29	Finance	Data Analyst	53	TRUE	4	Python	Good Culture	4	10	7	7	7	5	6.7
62a33d850f8c85	2022-10-06	08:48:00	00:01:06	Niger	Black or African	Unspecified	Male		22	Tech	Data Analyst	20	TRUE	3	Python	Better Salary	1	2	4	2	2	2	2.2
62a33d9bf8f756	2022-10-06	08:48:00	00:01:04	United States	Black or African	Unspecified	Male		23	Retail	Student/Looking	20	TRUE	4	Python	Better Salary	1	3	2	2	4	1	2.2
62a33dbdd8c00	2022-10-06	08:49:00	00:01:05	Canada	Asian or Asian A	Unspecified	Female		26	Tech	Student/Looking	53	FALSE	5	R	Better Salary	7	8	8	6	8	3	6.7
62a33de124a26	2022-10-06	08:49:00	00:01:21	Nigeria	Black or African	Unspecified	Male		28	Other	Data Analyst	20	FALSE	5	Python	Better Salary	1	1	3	1	2	0	1.0
62a33dedd8c00	2022-10-06	08:49:00	00:01:23	Kenya	Black or African	Unspecified	Male		29	Finance	Data Analyst	53	TRUE	2	Python	Remote Work	5	5	5	5	5	5	5.0
62a33dfb8dc00e	2022-10-06	08:50:00	00:01:35	United States	White or Caucas	Unspecified	Female		31	Tech	Data Analyst	115.5	TRUE	3	SQL	Remote Work	10	5	10	6	10	10	8.5
62a33e38f3072c	2022-10-06	08:51:00	00:01:25	Azerbaijan	White or Caucas	Unspecified	Male		28	Healthcare	Data Analyst	20	TRUE	3	Python	Remote Work	5	6	8	8	7	1	5.8
62a33e54d8c00	2022-10-06	08:51:00	00:01:47	Argentina	Hispanic or Latin	Unspecified	Male		32	Finance	Data Analyst	20	FALSE	3	Python	Better Salary	4	4	4	4	4	4	4.0
62a33e62575fa3	2022-10-06	08:51:00	00:02:14	Sudan	Black or African	Unspecified	Male		24	Tech	Data Analyst	20	FALSE	2	Python	Better Salary	7	8	8	8	3	8	7.0
62a33e63cec549	2022-10-06	08:51:00	00:02:21	India	Asian or Asian A	Unspecified	Male		31	Tech	Data Analyst	20	FALSE	2	Python	Good Work/Life	4	6	6	6	7	7	6.0
62a33e72bae91	2022-10-06	08:52:00	00:02:07	United States	White or Caucas	Unspecified	Male		23	Other	Data Analyst	75.5	FALSE	4	SQL	Responsibility	7	10	5	5	5	2	5.7
62a33e89f8c85	2022-10-06	08:52:00	00:01:20	United States	Asian or Asian A	Unspecified	Female		33	Energy	Data Analyst	115.5	TRUE	4	Python	Remote Work	1	6	9	1	0	2	3.2
62a33e8bf8d5ef	2022-10-06	08:52:00	00:01:34	United States	White or Caucas	Unspecified	Male		43	Healthcare	Data Analyst	95.5	FALSE	1	Other	Better Salary	6	7	7	7	7	4	6.3
62a33e90cec54f	2022-10-06	08:52:00	00:01:34	Nigeria	Other	Unspecified	Female		23	Energy	Student/Looking	20	TRUE	5	R	Remote Work	0	0	5	0	0	0	0.8
62a33e9df8f756	2022-10-06	08:52:00	00:01:47	United States	Black or African	Unspecified	Male		31	Finance	Student/Looking	53	TRUE	5	Python	Good Work/Life	10	7	10	10	5	3	7.5
62a33eb6bc686f	2022-10-06	08:53:00	00:01:38	United Kingdom	White or Caucas	Unspecified	Male		31	Finance	Data Analyst	75.5	TRUE	3	Python	Growth	5	6	7	5	4	4	5.2
62a33ee0bae91	2022-10-06	08:53:00	00:03:30	India	Asian or Asian A	Unspecified	Male		25	Other	Data Analyst	20	TRUE	4	Python	Good Work/Life	8	7	8	7	7	7	7.3
62a33ee80f8c85	2022-10-06	08:54:00	00:01:37	United States	White or Caucas	Unspecified	Male		43	Tech	Data Engineer	137.5	FALSE	2	SQL	Better Salary	8	10	10	9	9	9	9.2
62a33efed0a77c1	2022-10-06	08:54:00	00:02:08	Canada	Other	Unspecified	Female		26	Finance	Data Analyst	53	FALSE	4	Python	Better Salary	2	3	5	5	5	5	4.2
62a33f0ef408ba	2022-10-06	08:54:00	00:03:54	Japan	White or Caucas	Unspecified	Male		32	Automotive	Other	53	FALSE	3	Python	Remote Work	5	10	8	8	8	7	7.7
62a33f1f78f756f	2022-10-06	08:54:00	00:02:28	Greece	White or Caucas	Unspecified	Male		31	Tech	BI Analyst	20	FALSE	4	Python	Better Salary	2	9	7	7	5	6	6.0
62a33f52bc6861	2022-10-06	08:55:00	00:01:33	United States	White or Caucas	Unspecified	Male		29	Healthcare	Data Analyst	95.5	TRUE	3	SQL	Better Salary	3	5	6	4	4	5	4.5
62a33f56cf2555	2022-10-06	08:55:00	00:00:56	Peru	Hispanic or Latin	Unspecified	Male		24	Automotive	Data Analyst	20	TRUE	3	Python	Good Culture	3	1	3	2	3	1	2.2
62a33f5a5d257a	2022-10-06	08:55:00	00:02:38	United States	White or Caucas	Unspecified	Male		34	Construction	Student/Looking	53	TRUE	5	Python	Remote Work	2	0	3	2	0	0	1.0
62a33fa5cec549	2022-10-06	08:57:00	00:01:52	Australia	Other	Unspecified	Male		42	Other	Data Analyst	53	TRUE	5	Other	Better Salary	5	6	6	5	4	6	5.3
62a33fbcf575fa3	2022-10-06	08:57:00	00:02:03	United States	Asian or Asian A	Unspecified	Female		4	JavaScript	Student/Looking	20	TRUE	4	JavaScript	Remote Work	0	0	0	0	0	0	0.0
62a33fcf18134d	2022-10-06	08:57:00	00:02:32	Spain	White or Caucas	Unspecified	Male		43	Telecommunicat	Student/Looking	53	FALSE	2	Python	Good Work/Life	3	1	3	4	0	3	2.3
62a34006d8c00	2022-10-06	08:58:00	00:02:39	Kenya	Other	Unspecified	Male		23	Construction	Student/Looking	20	TRUE	5	Python	Better Salary	0	0	0	0	0	0	0.0
62a34020575fa3	2022-10-06	08:59:00	00:01:11	Japan	Asian or Asian A	Unspecified	Male		23	Education	Student/Looking	20	TRUE	4	Python	Better Salary	5	5	3	5	2	6	4.3
62a34038b4da2	2022-10-06	08:59:00	00:03:55	United States	White or Caucas	Unspecified	Male		28	Manufacturing	Data Analyst	95.5	TRUE	3	Python	Other	8	6	2	3	6	7	5.3
62a3406b575fa3	2022-10-06	09:00:00	00:01:25	United States	White or Caucas	Unspecified	Male		32	Finance	Data Analyst	75.5	FALSE	3	SQL	Better Salary	5	7	1	1	0	8	3.7
62a340cab4da2f	2022-10-06	09:02:00	00:02:47	United States	White or Caucas	Unspecified	Female		41	Manufacturing	Data Analyst	75.5	TRUE	4	R	Remote Work	7	8	10	8	9	9	8.5
62a340d62e54cf	2022-10-06	09:02:00	00:00:56	United Kingdom	White or Caucas	High School	Male		33	Finance	Data Analyst	53	TRUE	3	Python	Remote Work	5	6	6	6	6	6	5.8
62a340e10a77c1	2022-10-06	09:02:00	00:01:57	United States	Hispanic or Latin	Unspecified	Male		38	Tech	BI Analyst	187.5	TRUE	2	Python	Good Work/Life	10	8	9	8	8	10	8.8
62a340f5bc6861	2022-10-06	09:02:00	00:02:32	India	Other	Unspecified	Female		22	Tech	Data Analyst	20	FALSE	2	SQL	Better Salary	8	8	8	8	5	10	7.8
62a34117cec54f	2022-10-06	09:03:00	00:01:59	Barbados	Black or African	Bachelors	Male		36	Energy	Student/Looking	20	TRUE	4	R	Remote Work	1	3	6	6	6	9	5.2
62a3413a5d257f	2022-10-06	09:03:00	00:01:18	United States	Asian or Asian A	Masters	Female		28	Finance	Data Engineer	75.5	FALSE	2	Python	Remote Work	2	1	1	1	3	7	2.5
62a341b0f8f756	2022-10-06	09:05:00	00:01:44	United States	Hispanic or Latin	Bachelors	Female		24	Tech	Data Analyst	75.5	TRUE	4	Python	Other	6	6	6	6	5	6	5.8
62a341e8f8d5ef	2022-10-06	09:06:00	00:01:58	United States	White or Caucas	Bachelors	Male		29	Logistics	BI Analyst	53	FALSE	5	C/C++	Better Salary	3	6	7	7	0	3	4.3
62a34210cf2555	2022-10-06	09:07:00	00:02:15	Colombia	Hispanic or Latin	Associates	Male		29	Tech	Data Analyst	20	FALSE	2	Python	Better Salary	4	9	10	9	9	10	8.5
62a3424eb4da2f	2022-10-06	09:08:00	00:02:04	Germany	White or Caucas	Masters	Male		38	Automotive	BI Analyst	53	TRUE	3	Python	Good Work/Life	6	8	8	7	9	7	7.5
62a3425bbae91	2022-10-06	09:08:00	00:01:46	Bulgaria	White or Caucas	Masters	Male		31	Finance	Data Analyst	20	TRUE	4	N/A	Good Work/Life	3	5	4	3	3	3	3.5
62a3426cf408ba	2022-10-06	09:09:00	00:01:33	United States	White or Caucas	Masters	Male		25	Tech	Student/Looking	20	FALSE	4	Python	Better Salary	1	4	5	5	1	1	2.8
62a34272f8d5ef	2022-10-06	09:09:00	00:01:09	United States	White or Caucas	Masters	Male		24	Telecommunicat	Data Analyst	75.5	FALSE	3	Python	Better Salary	4	6	4	5	4	4	4.5
62a34340bc686f	2022-10-06	09:12:00	00:04:03	Turkey	White or Caucas	Bachelors	Male		28	Tech	BI Analyst	53	TRUE	4	Python	Better Salary	6	7	4	3	9	8	6.2
62a34383b4da2f	2022-10-06	09:13:00	00:01:03	India	Asian or Asian A	Bachelors	Female		23	Tech	Student/Looking	20	FALSE	3	Python	Good Culture	7	5	8	0	0	0	3.3
62a34389cec54f	2022-10-06	09:13:00	00:01:01	Spain	Hispanic or Latin	Masters	Male		21	Tech	Data Engineer	20	FALSE	3	R	Remote Work	2	3	2	5	5	7	4.0
62a343a22e54cf	2022-10-06	09:14:00	00:02:43	France	Other																		

Salary overview

Number of participants

610 (of 610)

Ethnographic filters

Country

All

Ethnicity

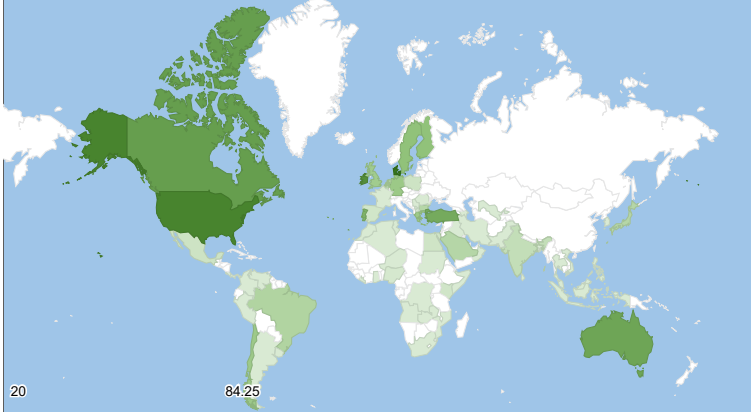
All

Age

All

Gender

All



Average salary

54 kUSD

Occupational filters

Current industry

All

Title

All

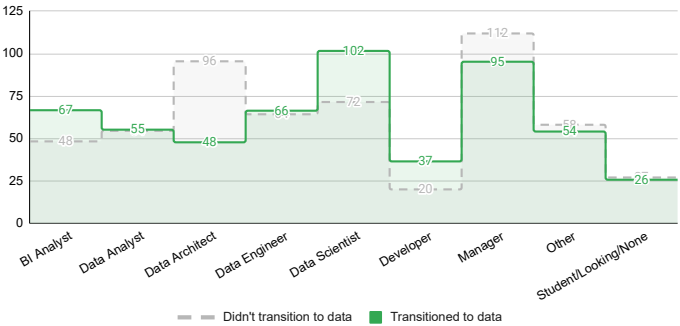
Has transitioned to data

All

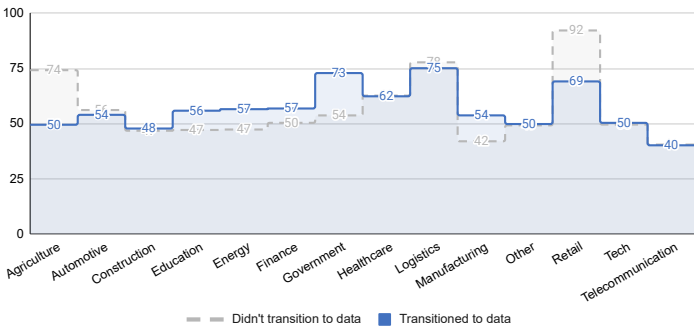
Top satisfaction metric

All

Salary gap by title



Salary gap by industry



Pivot data

By industry

AVERAGE of a switched_To_Data

industry_from	FALSE	TRUE
Agriculture	74	50
Automotive	56	54
Construction	47	48
Education	47	56
Energy	47	57
Finance	50	57
Government	54	73
Healthcare	63	62
Logistics	78	75
Manufacturing	42	54
Other	49	50
Retail	92	69
Tech	50	50
Telecommunication	41	40

By title

AVERAGE of a switched_To_Data

title	FALSE	TRUE
BI Analyst	48	67
Data Analyst	54	55
Data Architect	96	48
Data Engineer	64	66
Data Scientist	72	102
Developer	20	37
Manager	112	95
Other	58	54
Student/Lookin	27	26

Count of included participants, average salary

COUNTA of ID	AVERA
610	54

By country

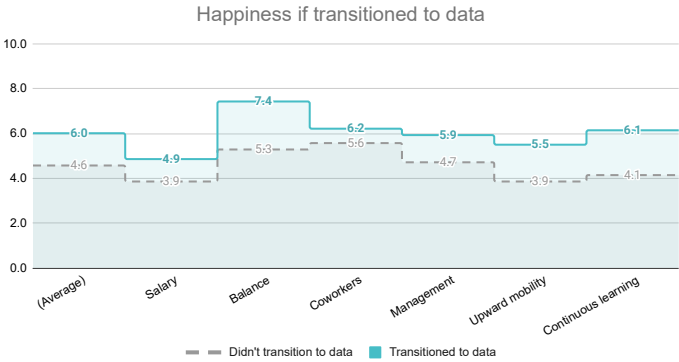
country	AVERA	COUNT
Algeria	20	1
Angola	20	1
Antigua	20	1
Argentina	20	7
Australia	65	8
Azerbaijan	20	1
Bangladesh	20	1
Barbados	20	1
Belgium	37	2

Happiness overview

Number of participants
21 (of 610)

Ethnographic filters

- Country 1 of 72
- Ethnicity All
- Age All
- Gender All

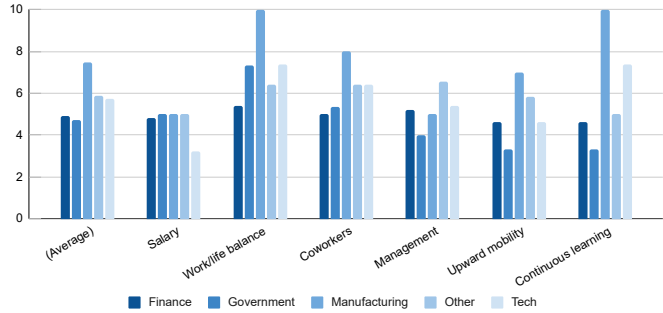


Overall happiness (out of 10)
5.5

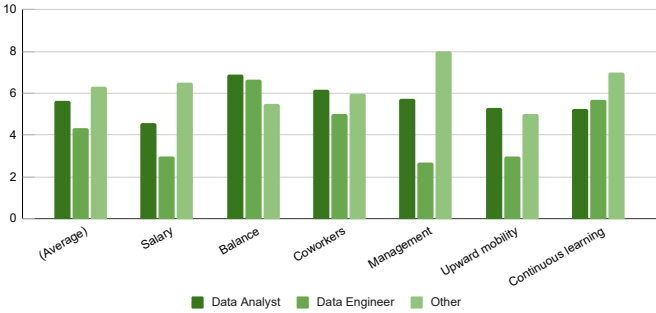
Occupational filters

- Current industry 5 of 10
- Title 3 of 4
- Has transitioned to data All
- Top satisfaction metric All

Happiness by current industry (categorized)



Happiness by title (categorized)



Pivot data

By industry

Industry	Average	Salary	Work/life	Coworkers	Management	Upward	Continuous
Finance	4.9	4.8	5.4	5.0	5.2	4.6	4.6
Government	4.7	5.0	7.3	5.3	4.0	3.3	3.3
Manufacturing	7.5	5.0	10.0	8.0	5.0	7.0	10.0
Other	5.9	5.0	6.4	6.4	6.6	5.9	5.0
Tech	5.7	3.2	7.4	6.4	5.4	4.6	7.4

By title

Title	Average	Salary	Balance	Coworkers	Management	Upward	Continuous
Data A	5.7	4.6	6.9	6.2	5.8	5.3	5.3
Data E	4.3	3.0	6.7	5.0	2.7	3.0	5.7
Other	6.3	6.5	5.5	6.0	8.0	5.0	7.0

By career switch

Switch	Average	Salary	Balance	Coworkers	Management	Upward	Continuous
FALSE	4.6	3.9	5.3	5.6	4.7	3.9	4.1
TRUE	6.0	4.9	7.4	6.2	5.9	5.5	6.1

Count of included participants, average happiness

Country	Average
21	5.5