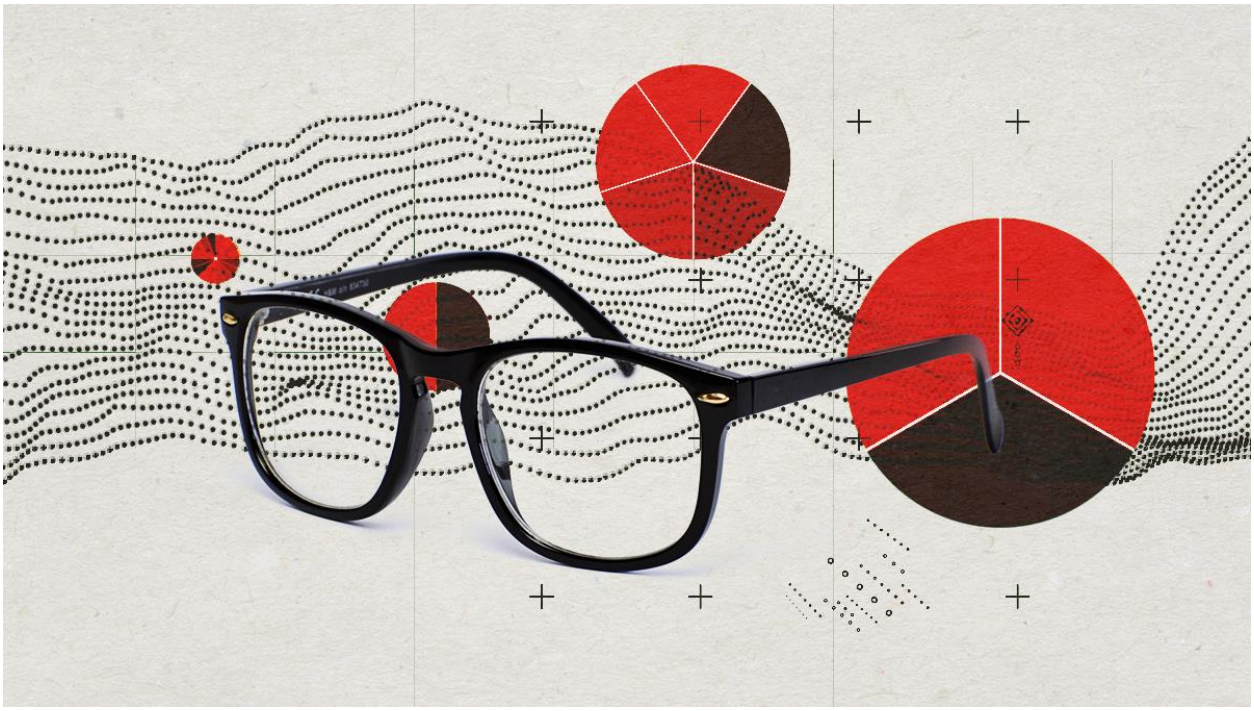
Advanced Analytics & Dashboard Design

Global Data Professional Salary

Summary

This dataset contains survey responses from 882 data professionals from 46 countries who took part in the 2019 Global Data Professional Salary Survey.



1HBR Staff/StudioM1/Moritz Otto/Getty Images

Data Source

The data is collected via surveys & provided by Brent Ozar in this link.

Data Contents

The dataset includes professional & personal details such as: location, salary, primary database, years of experience, education, hours of work per week, etc. from database administrators, data analysts, data architects, developers and data scientists in 2017-2019.

Data Limitations

- **Missing information on benefits and perks:** The dataset does not include details about non-monetary benefits or perks such as health insurance, vacation days, retirement plans, stock options, and others, which are important elements of a data professional's total compensation.
- **Absence of company-specific data:** The dataset does not provide sufficient information about the companies the data professionals work for, such as their industry or location. This information could provide valuable context for understanding salary variations.
- **Missing answers:** The surveyors gradually asked more questions over time, so if a question wasn't asked in a year, the answers are populated with *Not Asked*.
- **Unreliable information:** The postal code field was optional and may be unreliable.

Data Wrangling

Original column	Change	New column	Reason
None	**Created**	**id**	**Column contains unique identifiers**
Survey Year	Renamed	survey_year	Conformity to Python naming convention
Timestamp	Renamed	timestamp	Conformity to naming convention
SalaryUSD	Renamed	salary_in_usd	Conformity to naming convention
Country	Renamed	country	Conformity to naming convention
PostalCode	**Removed**	***None***	**Deemed unreliable by the author**
PrimaryDatabase	Renamed	primary_db	Conformity to naming convention
YearsWithThisDatabase	Renamed	years_of_exp_with_primary_db	Conformity to naming convention; Clarity
OtherDatabases	Renamed	other_dbs	Conformity to naming convention
EmploymentStatus	Renamed	employment_status	Conformity to naming convention
JobTitle	Renamed	job_title	Conformity to naming convention

ManageStaff	Renamed	manage_staff	Conformity to naming convention
YearsWithThisTypeOfJob	Renamed	years_of_exp_with_data_jobs	Conformity to naming convention; Clarity
HowManyCompanies	Renamed	number_of_companies_worked_for	Conformity to naming convention; Clarity
OtherPeopleOnYourTeam	Renamed	number_of_team_members	Conformity to naming convention; Clarity
CompanyEmployeesOverall	Renamed	number_of_company_employees	Conformity to naming convention; Clarity
DatabaseServers	Renamed	number_of_db_servers	Conformity to naming convention; Clarity
Education	Renamed	education	Conformity to naming convention
EducationIsComputerRelated	Renamed	education_is_computer_related	Conformity to naming convention
Certifications	Renamed	certifications	Conformity to naming convention
HoursWorkedPerWeek	Renamed	hours_worked_per_week	Conformity to naming convention
TelecommuteDaysPerWeek	Renamed	wfh_days_per_week	Conformity to naming convention; Clarity

PopulationOfLargestCityWithin20Miles	Renamed	pop_of_largest_city_within_20_miles	Conformity to naming convention
EmploymentSector	Renamed	employment_sector	Conformity to naming convention
LookingForAnotherJob	Renamed	looking_for_another_job	Conformity to naming convention
CareerPlansThisYear	Renamed	career_plans_this_year	Conformity to naming convention
Gender	Renamed	gender	Conformity to naming convention
OtherJobDuties	Renamed	other_job_duties	Conformity to naming convention
KindsOfTasksPerformed	Renamed	tasks_perfomed	Conformity to naming convention
Counter	**Removed**	***None***	**Column is redundant**

Data Consistency Checks

Column	Consistency issue	Possible reasons	Solution
id	*None*	*None*	*None*
survey_year	*None*	*None*	*None*
Timestamp	*None*	*None*	*None*

salary_in_usd	**Non-numeric values** **Special characters** **Abnormally high & low values**	**Incorrect input** **Hourly rates** **Part-time salaries** **Low salaries**	**Values below 1000 or above 1000000 are removed**
country	*None*	*None*	*None*
primary_db	*None*	*None*	*None*
years_of_exp_with_primary_db	**Abnormally high values**	**Incorrect input**	**Outliers are removed**
other_dbs	**Missing data**	**Respondents not sharing info**	**Missing values are replaced with "Not Provided"**
employment_status	*None*	*None*	*None*
job_title	*None*	*None*	*None*
manage_staff	*None*	*None*	*None*
years_of_exp_with_data_jobs	*None*	*None*	*None*
number_of_companies_worked_for	*None*	*None*	*None*
number_of_team_members	*None*	*None*	*None*

number_of_company_employees	*None*	*None*	*None*
number_of_db_servers	*None*	*None*	*None*
education	*None*	*None*	*None*
education_is_computer_related	**Missing data**	**Respondents not sharing info**	**Replaced with "Not Provided"**
certifications	*None*	*None*	*None*
hours_worked_per_week	Abnormally high values	Incorrect input	Outliers are removed
wfh_days_per_week	*None*	*None*	*None*
pop_of_largest_city_within_20_miles	*None*	*None*	*None*
employment_sector	*None*	*None*	*None*
looking_for_another_job	*None*	*None*	*None*
career_plans_this_year	*None*	*None*	*None*
gender	*None*	*None*	*None*
other_job_duties	**Missing data**	**Respondents not sharing info**	**Replaced with "Not Provided"**
tasks_perfomed	**Missing data**	**Respondents not sharing info**	**Replaced with "Not Provided"**

Column Details

Column	Column Description
id	Unique identifier
survey_year	The year the survey was conducted
Timestamp	Timestamp when the survey answer was submitted
salary_in_usd	Salary in USD
country	The country where the respondent lives
primary_db	The primary database that the respondent works with
years_of_exp_with_primary_db	The number of years of experience with the primary database
other_dbs	Other databases which the respondent also works with
employment_status	Employment status
job_title	Job title
manage_staff	Whether the respondent is working as a manager
years_of_exp_with_data_jobs	The number of years of experience with data-related jobs
number_of_companies_worked_for	The number of companies the respondent has worked for

number_of_team_members	The number of people working in the same team
number_of_company_employees	The number of people working in the same company
number_of_db_servers	The number of database servers that the respondents works with
education	Education level
education_is_computer_related	Whether the education is related to computer
certifications	Data-related certifications
hours_worked_per_week	Number of work hours per week
wfh_days_per_week	Number of work-from-home days per week
pop_of_largest_city_within_20_miles	The population of the largest city within 20 miles of the respondent's living location
employment_sector	The sector the respondent works in
looking_for_another_job	Whether the respondent is looking for as job
career_plans_this_year	Any career plans for the current year
other_job_duties	Any other job duties
tasks_perfomed	Any tasks performed at work

Summary Statistics

Variables	time -variant/-invariant	structured/unstructured	qualitative/quantitative	qualitative	quantitative
				nominal/ordinal/binary	discrete/continuous
id	time-invariant	structured	qualitative	nominal	
survey_year	time-invariant	structured	qualitative	ordinal	
salary_in_usd	time-variant	structured	quantitative		continuous
country	time-invariant	structured	qualitative	nominal	
primary_db	time-invariant	structured	qualitative	nominal	
years_of_exp_with_primary_db	time-invariant	structured	quantitative		continuous
other_dbs	time-invariant	unstructured	qualitative	nominal	
employment_status	time-invariant	structured	qualitative	nominal	
job_title	time-invariant	structured	qualitative	nominal	
manage_staff	time-invariant	structured	qualitative	binary	
years_of_exp_with_data_jobs	time-invariant	structured	quantitative		continuous

number_of_companier_worked_for	time-invariant	structured	qualitative	ordinal	
number_of_team_members	time-invariant	structured	qualitative	ordinal	
number_of_company_employees	time-invariant	structured	qualitative	ordinal	
number_of_db_servers	time-invariant	structured	quantitative		continuous
education	time-invariant	structured	qualitative	ordinal	
education_is_computer_related	time-invariant	structured	qualitative	nominal	
certifications	time-invariant	structured	qualitative	nominal	
hours_worked_per_week	time-invariant	structured	quantitative		continuous
wfh_days_per_week	time-invariant	structured	qualitative	ordinal	
pop_of_largest_city_within_20_miles	time-invariant	structured	qualitative	ordinal	
employment_sector	time-invariant	structured	qualitative	nominal	
looking_for_another_job	time-invariant	structured	qualitative	nominal	
career_plans_this_year	time-invariant	structured	qualitative	nominal	
other_job_duties	time-invariant	structured	qualitative	nominal	
tasks_perfomed	time-invariant	structured	qualitative	nominal	

Research Questions

- What is the median salary level by profession?
- What is the average income of data professionals in each country?
- How does experience level affect income?
- Are there salary gaps amongst data professionals between countries?