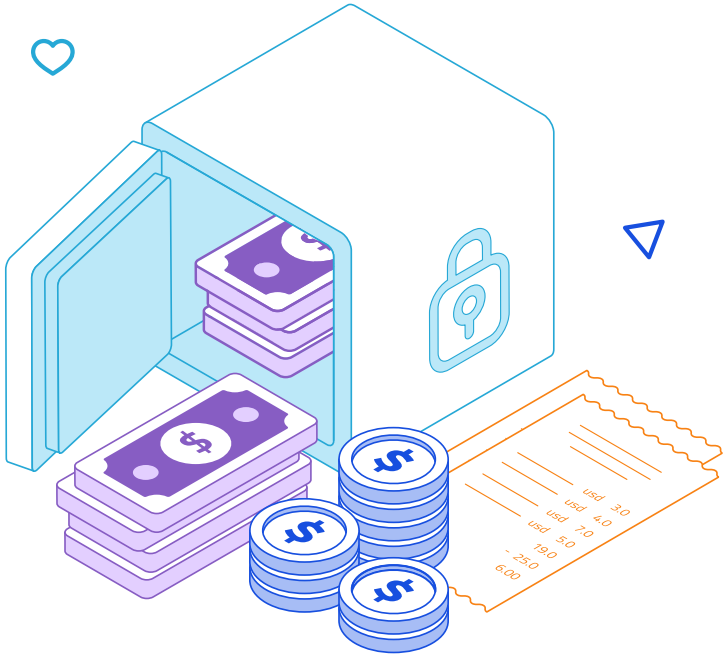
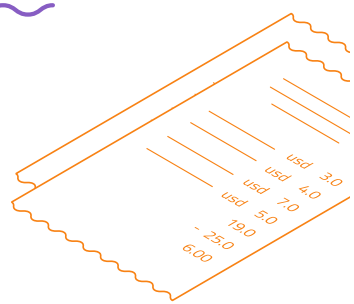


Data Science Salaries

Table of contents



Research Question

What technical skills would help to generate a higher pay for data analysts?

Why?

Important/Interesting

- Technical skills
- Varied Job Markets
- Roles and Responsibilities

Difficulty

- Experience and Seniority
- Data Availability
- Soft Skills

Dataset

Data Collection

- Kaggle
- Easy to get
- Scraped by the Kaggle user from “Glassdoor”
- Hard and impossible to collect info from former and current MSBA

Data Needed

- Need large data set >30 observations
- Cleaned and numerical

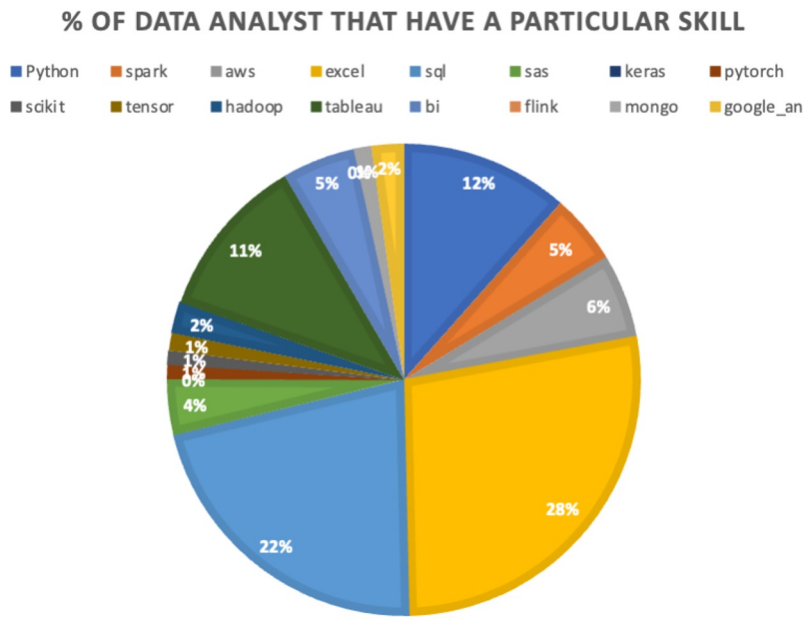
Data Set Desc.

- 2021 data analytic salaries
- 742 observations
- 42 variables but utilized the 16 variables

Statistical Analysis

Excel
28%

Python
22%



SQL
12%

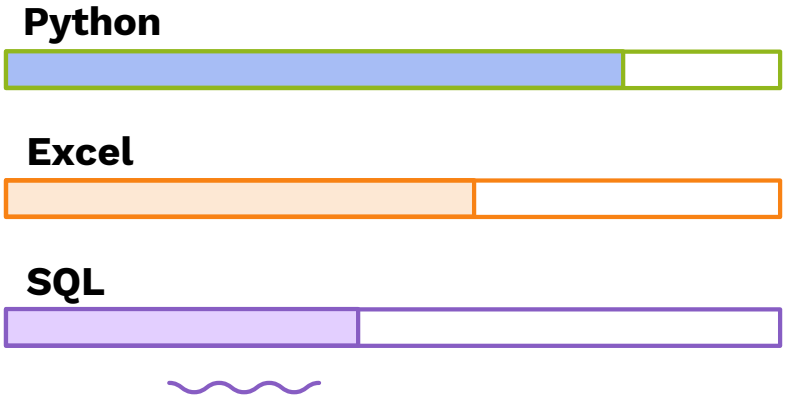
Tableau
11%

Preferred Skills by Sector

Popular Sectors

- Information Technology
- Business Services
- Insurance
- Healthcare
- Finance

Preferred Skills



Regression Analysis

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.430[a]	.185	.167	34.20965

a. Predictors: (Constant), Google=1.0, Scikit=1.0, Mango=1.0, Bi=1.0, Flink=1.0, Sas=1.0, Excel=0.0, Hadoop=1.0, AWS=1.0, Python=1.0, Pytorch=1.0, Tableau=1.0, Keras=1.0, SQL=1.0, Spark=1.0, Tenspr=1.0

b. Dependent Variable: AVG_Salary

Is there a relationship?

- Average income vs. Technical skills

Multiple regression Analysis

- R Square = 0.185
- A greater R Square would imply a stronger relationship.
- While R Square > O, is it statistically different than zero?

Hypothesis Testing

Hypothesis

- Ho: The regression model does not explain any of the total variation
- Ha: The regression model does explain any of the total variation
- $F = 10.285 > F_{0.05} = 1.5705$ | Reject Ho
- The sample data shows that the regression does explain any of the total variation at a 5% level of significance.

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	192588.486	16	12036.780	10.285	<.001[b]
	Residual	848467.585	725	1170.300		
	Total	1041056.072	741			

a. Dependent Variable: AVG_Salary

b. Predictors: (Constant), Google=1.0, Scikit=1.0, Mango=1.0, Bi=1.0, Flink=1.0, Sas=1.0, Excel=0.0, Hadoop=1.0, AWS=1.0, Python=1.0, Pytorch=1.0, Tableau=1.0, Keras=1.0, SQL=1.0, Spark=1.0, Tenspr=1.0

Conclusion

- **What we find**:
  - Python, Excel, SQL, Tableau are most popular and common skills for a Data analysts. Visualization tool such as Tableau and BI are also a popular skills in Healthcare, Finance industry
  - Skills does contribute to a higher pay
- **What is the limitation**:
  - Other than skills there are other factors that needs to be considered into a higher pay. such as where you work (Region), your level of experience, and the company size, etc.

Thank You!