

Analysis of Stack Overflow Survey 2017

Stack Overflow, a knowledge sharing platform that caters for programmers, conducts annual surveys of individuals who code worldwide across areas of interest such as programming languages, job satisfaction, work hours etc.

I obtained and analysed a dataset of responses from the 2017 survey in a bid to:

1. Determine the level of participation of respondents based in Africa. This subset of the dataset was obtained by extracting respondents who filled an African country in the survey.
2. Infer from the analysis, the distribution of programmers across the countries within Africa.
3. Identify common characteristics of African countries with the most respondents and
4. Establish that programmers based in certain African countries show higher career satisfaction than others outside these countries.

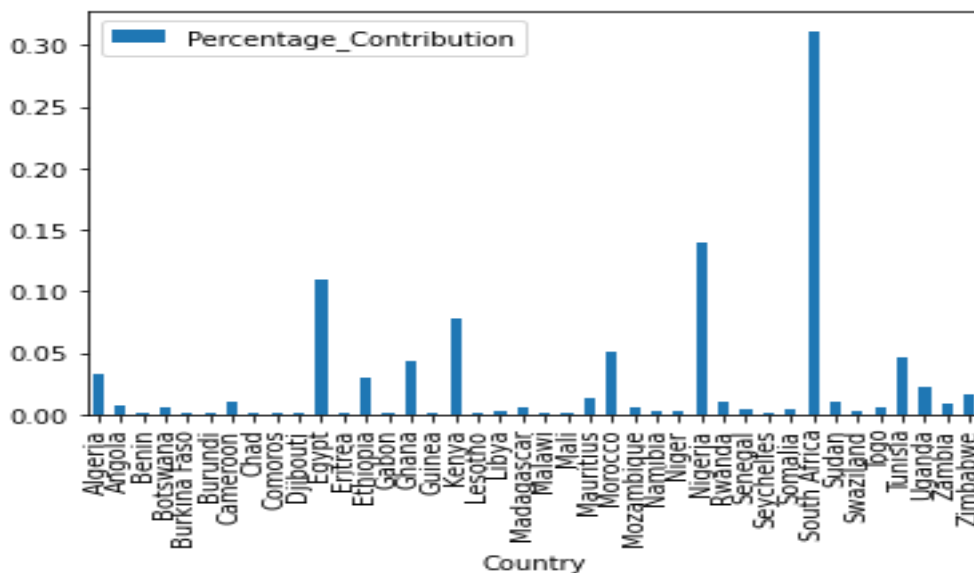
The dataset, sourced from Kaggle.com, was subjected to basic exploratory data analysis to determine the extent of missing values prior to use. The impact of incomplete records on the analysis was minimized by selecting relevant fields with nan records below a threshold. No records were dropped but missing numeric data within affected columns were filled using the column mean value.

1. What proportion of survey respondents identified with a country within Africa?

	Continent	Respondent	Percentage_Contribution
0	Africa	1221	2.375856
1	Outside Africa	50171	97.624144

2.38% of respondents identify with a country in Africa.

2. Which African countries have respondents on the survey?



Within Africa, the top 4 countries - South Africa, Nigeria, South Africa, South Africa - with over 63.9% of respondents are English speaking.

3. Given that top programming languages have English keywords/phrases, does the official language of the respondent's associated country have an impact on the respondent's career?

	Language	CareerSatisfaction
0	English	7.179658
1	Non English	6.889777

Career satisfaction appears higher in respondents from English-speaking countries compared to those from non-English speaking countries.

Conclusion

As seen in the analysis, the dataset can provide insights required to answer the identified business questions – participation level of programmers based in Africa, Geek Hotspots in Africa and likely impact of language on career satisfaction