

[Return to "Data Scientist Nanodegree" in the classroom](#)[DISCUSS ON STUDENT HUB](#)

Write a Data Science Blog Post

REVIEW

HISTORY

Requires Changes

2 SPECIFICATIONS REQUIRE CHANGES

Great job on very first submission and your project looks great because of your hard work 🙌🙌

Please look at all the tips and requirements, I shared to include in your project. KEEP UP THE GOOD WORK!

Don't get upset or disappointed, you did a great job which deserves a big compliment, think that those changes are a great opportunity to learn more and perfect your skills.

I wish you good luck with your Nanodegree. Looking forward to your next submission.

For any queries, you can ask on [Knowledge Portal](#) as well

Stay 🙌 ! Stay Safe

Code Functionality and Readability

✓ All the project code is contained in a Jupyter notebook, which demonstrates successful execution and output of the code.

All of the code runs and is functional.

✓ Code has easy-to-follow logical structure. The code uses comments effectively and/or Notebook Markdown cells correctly. The steps of the data science process (gather, assess, clean, analyze, model, visualize) are clearly identified with comments or Markdown cells, as well. The naming for variables and functions should be according to PEP8 style guide.

Your code is readable and modular. You have effectively used markdown cells to segment different sections in your notebook. This really helps to understand the flow of the analysis. Apart from that, I can clearly see that you have followed the pep8 style naming convention.



Each stage of the CRISP-DM process. These stages are:

1. Business Understanding
2. Data Understanding
3. Prepare Data
4. Data Modeling
5. Evaluate the Results

Business Understanding

In this notebook, I will be exploring the 2017 and 2018 Stack Overflow results in order to glean some information on professional developers and what to focus on to have the best career in India.

There will be three questions I will seek to answer in order to get this information:

1. What Programming Languages are most used to work and Which Programming Languages are most Required in Stackoverflow survey data of 2017 and 2018 ?
2. How does Programming Languages used at work relates with Programming Languages, People Wants to Learn in India According to Stackoverflow survey data of 2017 and 2018?
3. What are the most wanted Programming Languages in India According to Stackoverflow survey data of 2017 and 2018?

Data Understanding

In order to gain some understanding of the data. I have to do these steps:

You have done a nice work on documentation. Most people don't give emphasis on the this part. Keep it up.

Useful References:

[Jupyter Notebook Best Practices](#)

[Jupyter notebook shortcuts](#)

[PEP 8 -- Style Guide for Python Code](#)

🔄 Code is well documented and uses functions and classes as necessary. All functions include document strings. DRY principles are implemented.

As per udacity requirement, you need to provide `document string in all the functions.`

Prepare Data

```
In [70]: #First make copy of dataframe and then Filter a dataframe by country
def filter_country(df, column_filter, country, columns):
    df_copy = df
    for column in column_filter:
        df_copy = df_copy[df_copy[column_filter] == country].dropna(subset=[column])

    return df_copy

# Filtering the dataframe
india_2017 = filter_country(df_2017, 'Country', 'India', ['HaveWorkedLanguage', 'WantWorkLanguage'])
india_2018 = filter_country(df_2018, 'Country', 'India', ['LanguageWorkedWith', 'LanguageDesireNextYear'])

In [71]: #Splitting the columns
def split_column(df, column):
```

```
df_copy = df
columnSeries = df_copy[column].apply(lambda x: x.split(';'))
return columnSeries

# Splitting the dataframe by columns.
worked_languages_2017 = split_column(india_2017, 'HaveWorkedLanguage')
wanted_languages_2017 = split_column(india_2017, 'WantWorkLanguage')

worked_languages_2018 = split_column(india_2018, 'LanguageWorkedWith')
wanted_languages_2018 = split_column(india_2018, 'LanguageDesireNextYear')

In [72]: #Just Flating a nested list
def flat(array_list):
    object_list = []
    for row in array_list:
        for obj in row:
            object_list.append(obj.strip())

    return object_list

# Flating nested list objects.
list worked_languages_2017 = flat(worked_languages_2017)
list wanted_languages_2017 = flat(wanted_languages_2017)
```

Kindly provide the document string

This will help you and others to quickly grasp the understanding of the codes without getting diving deeper into codes.

Useful References:

[Python Docstrings](#)

[Docstring vs Comments](#)

[Do not Repeat Yourself](#)

Data

- ✓ Project follows the CRISP-DM process outlined for questions through communication. This can be done in the README or the notebook. If a question does not require machine learning, descriptive or inferential statistics should be used to create a compelling answer to a particular question.

This code does an exceptional job of following the CRISP-DM process. It was clear what part of the process you were on in each part of the notebook, which is difficult to do in a notebook. Nice job!

2. Data Understanding
3. Prepare Data
4. Data Modeling
5. Evaluate the Results

Business Understanding

In this notebook, I will be exploring the 2017 and 2018 Stack Overflow results in order to glean some information on professional developers and what to focus on to have the best career in India.

There will be three questions I will seek to answer in order to get this information:

1. What Programming Languages are most used to work and Which Programming Languages are most Required in Stackoverflow survey data of 2017 and 2018 ?
2. How does Programming Languages used at work relates with Programming Languages, People Who Want to Learn in India According to Stackoverflow survey data of 2017 and 2018?
3. What are the most wanted Programming Languages in India According to Stackoverflow survey data of 2017 and 2018?

Data Understanding

In order to gain some understanding of the data. I have to do these steps:

- Handle categorical and missing data
- Analyze, Model, and Visualize

```
In [132]: # import python libraries to handle datasets
```

Good Job

Useful References:

[Six steps in Crisp-dm process](#)

- 🔄 Categorical variables are handled appropriately for machine learning models (if models are created). Missing values are also handled appropriately for both descriptive and ML techniques. Document why a particular approach was used, and why it was appropriate for a particular situation.

Good job on handling the missing values.

But as per udacity requirement,

Document why a particular approach was used, and why it was appropriate for a particular situation.

Question 1: What Programming Languages are most used to work and Which Programming Languages are most Required in Stackoverflow survey data of 2017 and 2018 ?

Prepare Data

```
In [70]: #First make copy of dataframe and then Filter a dataframe by country
def filter_country(df, column_filter, country, columns):
    df_copy = df
    for column in columns:
        df_copy = df_copy[df_copy[column_filter] == country].dropna(subset=[column])
    return df_copy

# Filtering the dataframe
india_2017 = filter_country(df_2017, 'Country', 'India', ['HaveWorkedLanguage', 'WantWorkLanguage'])
india_2018 = filter_country(df_2018, 'Country', 'India', ['LanguageWorkedWith', 'LanguageDesireNextYear'])

In [71]: #Splitting the columns
def split_column(df, column):
    df_copy = df
    columnSeries = df_copy[column].apply(lambda x: x.split(';'))
    return columnSeries
```

Kindly elaborate, why you are dropping NaN values rather than dropping them.

Useful References:

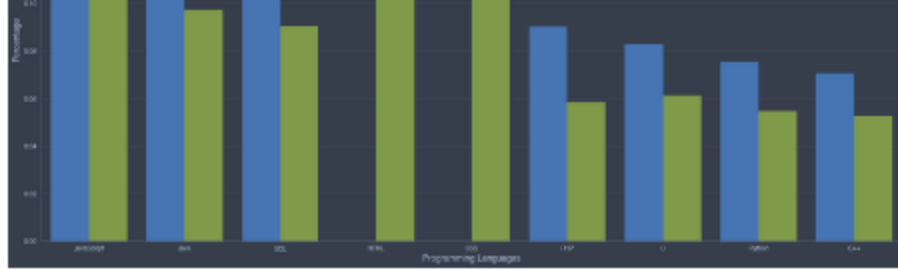
[Working with Missing Data](#)

Analysis, Modeling, Visualization

- ✓ There are between 3-5 questions asked, related to the business or real-world context of the data. Each question is answered with an appropriate visualization, table, or statistic.

Nice job with coming with your business questions. You have some really interesting questions with some cool findings. You have nicely put forward your hypotheses with support of great visual. Well Done!





Evaluate the Results

- The Highest Rate of percentage among all Programming Languages JavaScript have highest growth rate in 2017 which is around 16%, whereas this percentage is drop significantly in 2018 about 4% in India.
- As we can see through Graph, Java has same drop percentage about 4% in India.
- The Most Interesting this about this Graph is HTML and CSS it was probably not possible to select these two Languages as option in 2017 whereas in 2018 these two Programming Languages have risen such as HTML and CSS.

Question 2: How does Programming Languages used at work relates with Programming Languages, People Wants to Learn in India According to Stackoverflow survey data of 2017 and 2018?



Github Repository

- ✓ Student must have a Github repository of their project. The repository must have a README.md file that communicates the libraries used, the motivation for the project, the files in the repository with a small description of each, a summary of the results of the analysis, and necessary acknowledgements. Students should not use another student's code to complete the project, but they may use other references on the web including StackOverflow and Kaggle to complete the project.

Nice job posting your code to Github. Your README looks great! Your Table of Contents clearly lays out the installation, project motivation, file descriptions, results, and acknowledgements.

Useful References:

[Manage your data science project structure in early stage](#)

[How to write a good readme for your github project?](#)

Blog Post

- ✓ Student must have a blog post on a platform of their own choice (can be on their website, a Medium post or Github blog post). Student must communicate their results clearly. The post should not dive into technical details or difficulties of the analysis - this should be saved for Github. The post should be understandable for non-technical people from many fields.

Awesome work on the blog post.

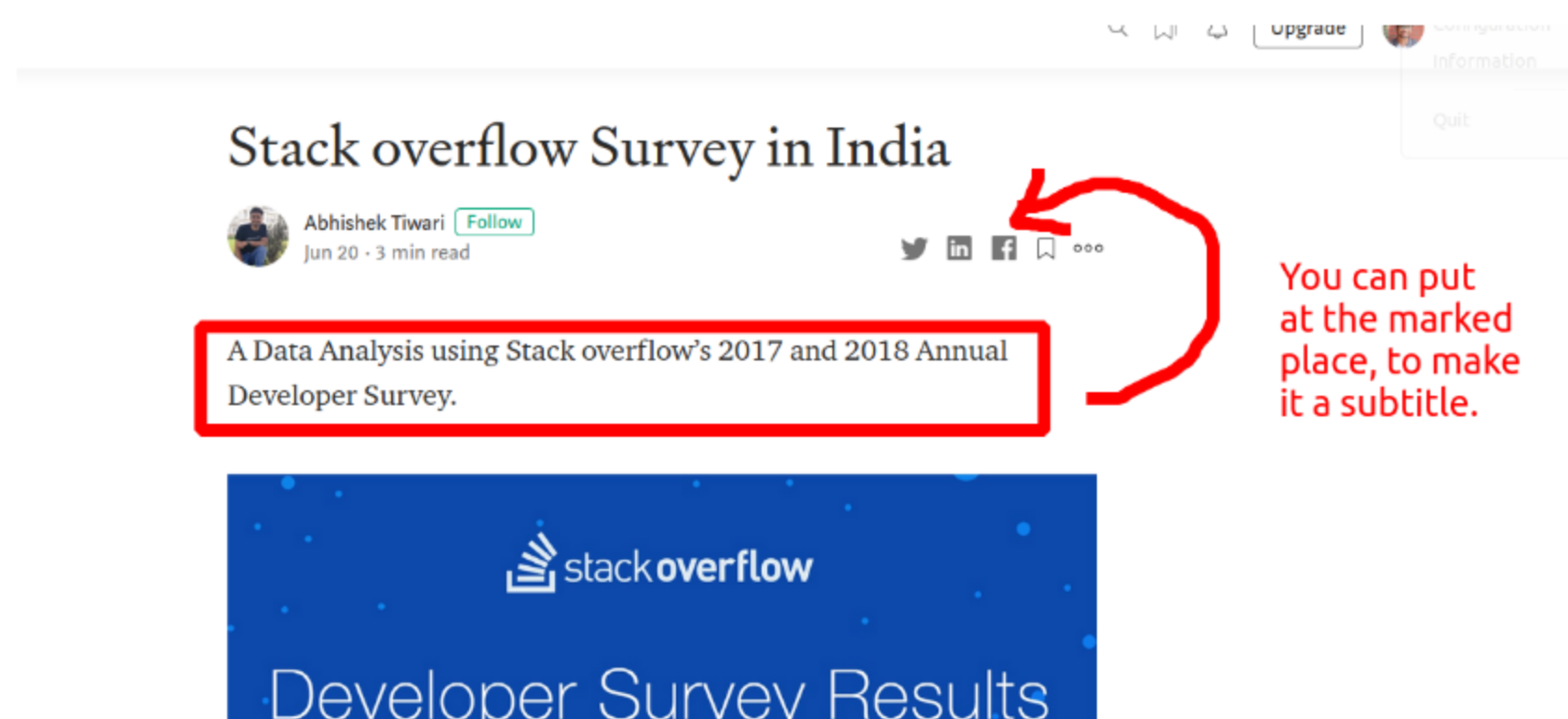
I really liked how you structure your analysis. Your post clearly communicates findings to non-technical stakeholders. You did a good job of providing a clear structure from the intro, the questions of interest, and a conclusion.

Useful References:

[Art of Storytelling](#)

- ✓ Student must have a title and image to draw readers to their post.

Nice job with the title and image. I definitely would click!



Useful References:

[Medium Titles, Subtitles, and Kickers](#)

[How you write headlines](#)

- ✓ There are no long, ongoing blocks of text without line breaks or images for separation anywhere in the post.

You did a great job of making sure your post was broken into short paragraphs with clear ideas in each.

- ✓ Each question is answered with a clear visual, table, or statistic that provides how the data supports or disagrees with some hypothesis that could be formed by each question of interest.

Each question was followed with a visual of what provided a potential solution to the question. The discussion around these plots is really well done too!

RESUBMIT PROJECT

