

Methods of Data Collection

– Sohail Hussain

Master of Computer Applications (MCA),RV College of Engineering, Bangalore, karnataka

 <https://orcid.org/0000-0002-2569-6927>  sohailhussain5@gmail.com

Article History

Paper Nomenclature:

Research Reflection (RR)

Paper Code: CYBNMV1N6NOV2019RR1

Submission Online: 02-Nov-2019

Manuscript Acknowledged: 11-Nov-2019

Originality Check: 13-Nov-2019

Originality Test Ratio: 78% (Disclaimer)

Peer Reviewers Comment: 15-Nov-2019

Blind Reviewers Remarks 20-Nov-2019

Author Revert: 21-Nov-2019

Camera-Ready-Copy: 23-Nov-2019

Editorial Board Citation: 24-Nov-2019

Published Online First: 22-Jan-2020

With the increasing number of security threats and data breaches it's very necessary to set up a security solution for any system. One can either pay someone to do it or do it himself. Both ways there is always a chance of missing out on something unless we have a perfect guide to tick the completed work. This article intends to do exactly that by putting light on all the factors which are needed to be considered while setting up any system. One may think that this is purely for business purposes but even personal rigs will benefit from this. This is because personal rigs are the ones who have users of different ages and understanding. From where to start and what to consider to what to choose and how much to pay (or do you need to pay) is what the article intends to deliver.

Keywords : Innovation | Data Management | Govt Portals Explore | Analyse | Share Quality Data

Kaggle

It supports a variety of dataset publication formats, but we strongly encourage dataset publishers to share their data in an accessible, non-proprietary format if possible. Not only are open, accessible data formats better supported on the platform, they are also easier to work with for more people regardless of their tools.

Kaggle supports database files using the lightweight SQLite format. SQLite databases consist of multiple tables, each of which contains data in tabular format. These tables support large datasets better than CSV files do, but are otherwise similar in practice.

Data.gov.in

Open Government Data (OGD) Platform India - data.gov.in - is a platform for supporting Open Data initiative of Government of India. The portal is intended to be used by Government of India Ministries/ Departments their organizations

to publish datasets, documents, services, tools and applications collected by them for public use. It intends to increase transparency in the functioning of Government and also open avenues for many more innovative uses of Government Data to give different perspective. The base Open Government Data Platform India is a joint initiative of Government of India and US Government. Open Government Data Platform India is also packaged as a product and made available in open source for implementation by countries globally.

dgft.gov.in

Directorate General of Foreign Trade (DGFT) organisation is an attached office of the Ministry of Commerce and Industry and is headed by Director General of Foreign Trade. Right from its inception till 1991, when liberalization in the economic policies of the Government took place, this organization has been essentially involved in the regulation

and promotion of foreign trade through regulation. Keeping in line with liberalization and globalization and the overall objective of increasing of exports, DGFT has since been assigned the role of "facilitator". The shift was from prohibition and control of imports/exports to promotion and facilitation of exports/imports, keeping in view the interests of the country.

commerce.gov.in

The Department formulates, implements and monitors the Foreign Trade Policy (FTP) which provides the basic framework of policy and strategy to be followed for promoting exports and trade. The Trade Policy is periodically reviewed to incorporate changes necessary to take care of emerging economic scenarios both in the domestic and international economy

The Department is functionally organized into the following 10 Divisions:

1. International Trade Policy Divisions
2. Foreign Trade Territorial Division
3. Export Products Division
4. Export Industries Division
5. Export Services Division
6. Economic Division
7. Administration & General Service Division
8. Finance Division
9. Supply Division
10. Logistics Division

The Three tools that are most popular with data analysts, Excel, R, Python,

1. Excel

- Data management and storage of datasets.
- Simple statistical analysis for regression analysis.
- Combine Word and PowerPoint to create data analysis reports.
- Produces charts for data visualization.

1.2 Advantages

- It's easy to get started with Excel.
- The learning resources are very rich.
- You can do a lot of things with Excel: modelling, visualization, reports, dynamic charts, etc.

1.3 Disadvantages

To fully master Excel, you need to learn VBA (**Visual Basic for Applications**) so the difficulty is still very high.

- When the amount of data is large, there will be a situation of data redundancy.

- The Excel data file itself can hold only 1.08 million rows without the aid of other tools, and it's not suitable for processing large-scale data sets.
- The built-in statistical analysis is too simple and has little practical value.
- Unlike Python, R, and other open source software, there is a charge for the genuine Excel.

2. R Studio

The functions of R cover almost any area where data is needed. As far as our general data analysis or academic data analysis work is concerned, the things that R can do mainly include the following aspects.

- Data cleaning and data reduction.
- Web crawling.
- Data visualization.
- Statistical hypothesis testing (t test, analysis of variance, chi-square test, etc.).
- Statistical modelling (linear regression, logistic regression, tree model, neural network, etc.).
- Data analysis report output (R markdown).

3. Python Pandas

- Data crawling.
- Data cleaning.
- Data modelling.
- Construct data analysis algorithms based on the business scenarios and actual problems.
- Data visualization.
- Advanced fields of data mining and analysis, such as machine learning and text mining.

Comparison between R and Python

R and Python are both data analysis tools that need to be programmed. The difference is that R is used exclusively in the field of data analysis, while scientific computing and data analysis are just an application branch of Python. Python can also be used to develop web pages, develop games, develop system backends, and do some operation and maintenance work.

A current trend is that Python is catching up with R in the field of data analysis. In some respects, it has surpassed R, such as machine learning and text mining. But R still maintains an advantage in the field of statistics. The development of Python in data analysis has modelled some of the features of R in many places.

If you can only choose one of them to learn because of the limited time, I recommend using Python. Python mimics many features of R, such as DataFrames in the Pandas library. And the visualization package under development, ggplot, mimics the very famous ggplot2 in R.

Implementation

Data visualization is required for data analysis because the text is not as good as the table, and the table is not as good as the graph. The charts of Excel can meet basic graphics requirements, but this is only the basis. The advanced visualizations require programming in programming languages such as R and Python.



Sohail Hussain is a Master of Computer Applications (MCA) from RV College of Engineering, Bangalore, Karnataka and keen in working in IT & Online Services.

 sohailhussain5@gmail.com

Annexure I

Submission Date	Submission Id	Word Count	Character Count
13-Nov-2019	D62282038 (urkund)	1150	6327



Urkund Analysis Result

Analysed Document: 107-Research Instrument-230-1-2-20200113 (1).docx (D62282038)
Submitted: 11/11/2019 2:21:00 PM
Submitted By: skesharwani@ignou.ac.in
Significance: 78 %

Sources included in the report:

- <https://www.asiaportal.info/database/open-government-data-ogd-platform-india/>
- <https://www.civilserviceindia.com/Indian-Trade-Service.html>
- <https://www.jindaramexports.com/partners.html>
- <https://towardsdatascience.com/comparison-of-data-analysis-tools-excel-r-python-and-bi-tools-6c4685a8ea6f>

Instances where selected sources appear: 15

Note: Cybernomics runs an Urkund plagiarism tool for the originality check of an article before publication. Urkund is developed by Prio Infocenter AB based in Stockholm, Sweden.

Disclaimer: All Views expressed in this paper are my own, which some of the content are taken from open source website for the knowledge purpose. Those some of i had mentioned above in references section.

Reviewers Comment

Reviewer Comment 1: We have Many software tools for data management, and these are called Clinical Data Management Systems (CDMS).

Reviewer Comment 2: Python pandas is an open source, BSD-licensed library for providing high-performance, easy-to-use data structures, data analysis tools for the programming language.

Reviewer Comment 3: The author has well explained all the issue on data security and data management technique.

Editorial Excerpt

This article has 78% plagiarism which was accepted my taking disclaimer from author. This article is analysis the various concept of data security , data management , and some important government portals like dgft.gov.in , commerce.gov.in , Data.gov.in, and their work, advantage , and disadvantage, The Python Pandas Data management tools is very essential or each and every database world.,Hence this article is marked under "Research Instrument " category.

Citation

Sohail Hussain
 "Methods of Data Collection"
 Volume-1, Issue-6, Nov 2019. (www.cybernomics.in)

Frequency: Monthly, Published: 2019
 Conflict of Interest: Author of a Paper had no conflict neither financially nor academically.

