Analysing Implementation of Data Journalism in English Newspapers: Proposing Models for Data-Driven Reporting in Newspapers

Jenitta Sabu [1] & Devjani Talukdar [2]

Abstract

In the year 2014, data journalism as a field emerged globally and opened new doors for research in media studies. Researchers introduced various process models of data journalism including The Chicago Tribune Model (2013) and The Workflow Model of Data Journalism (2017). Further The Taxonomy Model of Data Journalism (2017) was introduced to give a new dimension to data journalism projects. However, this model emphasized on classification of data journalism projects in online platforms unlike static mediums of communication namely, newspaper and television.

The aim of this paper was to investigate whether the Taxonomy Model of Data Journalism will be suitable for newspapers and if not can there be any suitable modifications that can be proposed as extended model of data journalism especially for data-driven news stories in the newspapers. This research was based on quantitative approach using content analysis research technique and took into consideration secondary tool of data collection. The researcher made use of unstructured sampling for the study which included 3 national English daily newspapers of India namely, Times of India, Hindustan Times, and The Hindu. The sample size was newspapers of 10 consecutive days of January 2020. As per the results of the study, researcher has proposed two extended model of data-driven reporting in newspapers which will help print journalists to follow the more structuralized model of data journalism and will open a new chapter in educational strategies of data journalism.

Keywords: Data Journalism, Models of Data-Driven Reporting in Newspapers, Content Analysis, Print Journalism and Educational Strategies for Data Journalism

[1] Former Student, DME Media School, Delhi Metropolitan Education, Jenittasabu18@gmail.com
[2] Executive, News and Guest Coordination, Republic Media Network, devjanit@gmail.com

1. Introduction

The American Press Institute defined data journalism as gathering, cleaning, organizing, analyzing, visualizing, and publishing data to support the creation of a set of journalism. Since the year 1821 when the first example of data journalism was visible in The Guardian, data journalism as a field has emerged a lot and reached its peak in the year 2014. Hence the field of data journalism has opened new doors to media research. Contributing to the field of data journalism researchers introduced various process models of data journalism including The Chicago Tribune Model (2013) and The Workflow Model of Data Journalism (2017). Further, The Taxonomy Model of Data Journalism (2017) was introduced to give a new dimension to data journalism projects. However, this model emphasized the classification of data journalism projects in online platforms unlike static mediums of communication namely newspaper and television.

This study analyzed 240 data-driven news stories from three national English dailies to investigate whether 'The Taxonomy Model of Data Journalism-2017 introduced by Andreas Veglis and Charalampos Bratsas applied to data journalism practice in newspapers or not. If not, the researcher wanted to figure out if any modifications can be suggested as an extended model of data journalism.

In the end, as per the results of the study, the researcher proposed two extended models of data-driven reporting in newspapers namely, '*Model of data-driven reporting in newspapers based on purpose and relationship' and 'Model of data-driven reporting in newspapers based on the classification of data presentation, type, and structure'.* These proposed models will help print journalists to follow the more structuralized model of data journalism and will open a new chapter in educational strategies of data journalism.

2. Objectives

This research paper focused on analyzing the coverage of data-driven news stories in three main daily newspapers of India namely: The Hindu, Times of India, and Hindustan Times. This study is based on the reference of 'The Taxonomy Model of Data Journalism-2017' proposed by Andreas Veglis and Charalampos Bratsas in their paper titled 'Towards a Taxonomy Model of Data Journalism'. The main aim of this paper is to investigate whether the Taxonomy Model of Data Journalism will be suitable for the newspapers and if not can

there be any suitable modifications that can be proposed as an extended model of data journalism especially for data-driven news stories in the newspapers.

3. Literature Review:

The literature review section of this study was divided into 5 subparts: (a) History of data journalism, (b) The emergence and impact of data journalism in the media organizations, (c) Educational importance and ethics of data journalism, (d) Process and models of data journalism and (e) Critical review on the practice of data journalism.

a. History of data journalism: Research study titled 'The Art and Science of Data Journalism' (Howard, B) focused briefly on the rise of data journalism and the emphasis on The Guardian which turned out to be the first media organization to carry out the process. The paper also elaborated on the significance of the year 2014 which turned out to be a boom period for data journalism with the re-launch of Nate Silver's FiveThirtyEight.com, and Vox Media's general news site Vox.com as well as news ventures from the New York Times and Washington Post. The research project 'Teaching Data and Computational Journalism' (Berret, C. Phillips, C. 2016) accounted for the rise of data journalism from the year 1967 to 2013. The study emphasized the contributions of Philip Meyer. The study logged major developments in the practice of data journalism across the globe.

b. The emergence and impact of data journalism on the media organizations: Research paper titled 'Adapting investigative and data journalism to new players: How are actors working globally?' focused on media organizations coping with the changes of the digital age. The methodology used for research was personal interviews; the researcher took in-depth interviews of six professionals working within third party organizations in Asia, Africa, America, Europe, and Oceania. The results showed that number of news outlets stating that services have been reduced as many organizations and newsrooms are opting for advanced technologies. The results of the study were limited to six organizations and were based on a qualitative approach hence the sample size is not enough to generalize a conclusion. Research project titled 'Data Journalism in New Media Firms- the Role of Information Technology to Master Challenges and Embrace Opportunities of Data-Driven Journalism Projects' (Dal Zotto C; Schenker Y; Lugmayr A, 2015) stated that data journalism has 3 main dimensions and requires 3types of journalistic skills like computer-assisted reporting, news application development, and data visualizations. The study was formatted based on a review article that derived its conclusion from the literature review, experiences of practitioners, and analysis of

the best practices in the field of data journalism. The study mainly emphasized on industries collaborating with information technology and journalistic practices to widen the horizon of data journalism across the globe. However, the study was mentioned to be in progress as researches wanted to make use of a quantitative approach to understand how the collaboration will encourage the growth of data journalism.

c.   Educational importance and ethics of data journalism: In the research report titled 'Teaching Data and Computational Journalism', published by Colombia Journalism School and Knight Foundation, New York, the main objective of the researchers was to augment journalism with data-driven and computational techniques. The report emphasized the state of data journalism education, to underline the urgency of incorporating these skills to equip by the next generation of reporters and to offer guidelines to the budding journalists to make their further move in this direction. Hence the researchers proposed a curriculum model for data journalism which included- Integrating data as a core course (foundations of data journalism), Integrating data and computation to existing courses and concentrations, and Concentration in data and computation. This research report focused on entirely different aspects of data journalism that is the educational requirement of establishing data journalism as compulsory coursework in the discipline of Journalism and Mass Communication. The study also suggested required coursework that should be taken into consideration to prepare budding journalists in the age of the digital era. However, the sample taken into consideration was the 63 syllabi of colleges with the permission of government bodies, thus it would have been more useful to know whether students had any idea of data journalism and computation especially when only 14% of the classes focused on advanced classes. This data would have supported the research and automatically will arise as a need at the institutional level to introduce data journalism as compulsory coursework for a budding journalist.

'The Ethics of Data Journalism' published in the professional projects from the College of Journalism and Mass Communication, University of Nebraska, Lincoln focused on highlighting key ethical guidelines which included the importance of placing all data in context and minimizing the harm to the news. As data journalism demands special considerations, many of the ethical guidelines that are needed to be practiced are already in place according to the researcher. The paper investigates the ethics of data journalism in the context of education and made use of personal interviews with the industry expert and analysis of the latest research works along with best practices in the field of data journalism (McBride R, 2016). This study was a review article that was based on in-depth interviews of the professionals, analysis of the latest research findings, and best practices in data journalism. Hence these ethical guidelines

are not backed by quantitative data analysis to support its generalization to every news organization and data journalist. Hence, this provides a scope of research for other researchers to figure out whether data journalists adhere to these guidelines, or do they differ in practice.

d. Process and models of data journalism: The research paper titled Data Journalism: An outlook of the future process described the Data Journalism Process Model of The Chicago Tribune-2013 in detail. The Chicago Tribune has a specialized data journalism team that has their-own Maps and Apps. Unlike many teams in this field, the news app of Chicago Tribune is a bit special group of hackers that was founded by computer technicians for whom journalism was a significant career change. Some of the professionals have acquired a master's degree in journalism after several years of coding for business purposes, whereas as some worked for open government communities. The Chicago Tribune Model has nine processes that are compiled into four main phases which include finding the story lead, data manipulation, story creation, and story visualization (Rapeli M, 2013). For this study, 3 professionals were interviewed to determine the authenticity of the future streamlined model of the Chicago Tribune. The three professionals had been working in the field of data journalism for two or more years and had experience in working with the major Finnish news organizations. One interviewee worked as a data journalism producer in Helsingin Sanomat, one of the leading newspapers of Finland, rest two of them worked as data journalism producers and programmers in Finnish Broadcasting Company. However, two interviewees from the same organization will have similar results to share with the researcher as the working patterns remain the same for the organization. The results of the interview could have been different if colleagues of three different organizations were interviewed which would have helped in generalizing the future streamlined model of Chicago Tribune especially in media organizations of Finland.

The authors of the research paper titled 'Reporters in the age of data journalism' proposed the workflow model of data journalism in the year 2017. The model suggested broader steps in data journalism, which a journalist must practice for assuring the authenticity of the news. The study used a survey research design and used online questionnaires as the data collection tool. The questionnaires covered how journalists use the data, level of expertise, the learning needs, and the barriers they face and the barriers they face. The questionnaires were distributed through the School of Journalism and Mass Communication-Aristotle University. The participants of the research were the members of the Journalist's Union of Macedonia, Thrace Daily Newspaper, and Open Knowledge Foundation (Veglis A; Bratsas C, 2017). This research study opted for an online questionnaire tool for data collection, thus the results can be prone to biases that might have affected the study. Further due to online tool of data collection

there was less control over the respondents hence this might have led to low responsiveness of the respondents as there were just 58 participants for this research which is not presumed by most of the statisticians to be the minimum standard size of a sample which is 100, whereas for descriptive research 30 subjects are suitable.

'Towards A Taxonomy of Data Journalism' focused on proposing a taxonomy of data journalism projects that can help the future journalists to choose the appropriate type of data project which will be suitable for their needs. The proposed taxonomy took into account various parameters that play an important role in data journalism projects with special emphasis on the type and interactivity role of data visualization (Veglis A; Bratsas C, 2017). The implementation of this model was further analyzed in the research paper titled 'Visualization and Interactivity in Data Journalism Projects' authored by Christina Karypidou, Charalampos Bratsas, and Andreas Veglis, the researchers took into account the proposed taxonomy model of data journalism to determine the visualization effect and level of interactivity among the data journalism projects. Authors analyzed the websites of The Guardian and The New York Times - categorized as the elite newsrooms of the world, The British Broadcasting Corporation, and The Cable News Network- both of them are two most famous networks and the two largest news agencies namely, Associated Press and Reuters. Hence 4 major organizations of media industry were taken into consideration to check the implementation of the model and most important in both the papers the researchers analyzed the data-driven coverage in the online portals to determine the authenticity of the model.

e.   A critical review on the practice of data journalism: In the research paper 'Visualization and Interactivity in Data Journalism Projects' (Veglis A; Bratsas C; Karypidou C, 2019) researchers took into account the proposed taxonomy model of data journalism to determine the visualization effect and level of interactivity among the data journalism projects. Authors analyzed the websites of The Guardian and The New York Times - categorized as the elite newsrooms of the world, The British Broadcasting Corporation, and The Cable News Network- both of them are two most famous networks and the two largest news agencies namely, Associated Press and Reuters. The paper stated that the emergence of data journalism formed new conditions for writing and distributing the news. Professionals adopt new methods and the public makes use of new practices on how to read the news. More and more organizations are using data journalism to tell their stories. It is thus believed that data journalism is one of the pillars of modern journalism. In the era of big data, data journalism can play a significant role in utilizing the available data sets. The paper discussed the issues of visualization and interactivity levels in data journalism projects. In this research paper, it was proved that the

taxonomy model could be used in all types of data journalism projects. In the future, the effectiveness of the proposed model could be extended by including other types of visualizations that can be categorized into data-rich visualizations to communicate stories with complex data in an engaging manner.

The research project titled 'Data Journalism Concept and Practices in India: A case study of data journalism initiative in India' by Indiaspend.com focused on the development of data journalism in India and emphasized on the concept and practices of data journalism in the Indian context. The websites of print media namely-The Hindu, The Indian Express, and The Times of India were analyzed based on the following parameters which were divided into sections of website and content. The parameters of the website included a section of the website and frequency of updating, whereas the parameters of the content included the nature of the story, arrangement of the story, source of data, and the number of graphs and tables.

In both the research papers, the researchers have analyzed the online versions of major media entities for example researchers Christina Karypidou, Charalampos Bratsas, and Andreas Veglis took into consideration the online portals of The Guardian and The New York Times - categorized as the elite newsrooms of the world, The British Broadcasting Corporation, and The Cable News Network- both of them are two most famous networks and the two largest news agencies namely, Associated Press and Reuters, to analyze the implementation of their model associated with data journalism projects. Similarly in India, Indiapend.com focused on the development of data journalism in India and emphasized on the concept and practices of data journalism in the Indian context. The websites of print media namely-The Hindu, The Indian Express, and The Times of India were analyzed based on the following parameters which were divided into sections of website and content.

In both, the case analysis of data journalism was limited to the online portals of media organizations. However, in this study, the researcher focused on analyzing the coverage of data journalism in three main daily papers of India namely, The Hindu, The Times of India and Hindustan Times (as per the Indian Readership Survey) based on the taxonomy model of data journalism-2017. The main aim of this paper was to investigate if the model was suitable for data-driven coverage in the newspapers and if not were there any modifications that can be proposed as an extended model of this taxonomy based on the practice of data journalism in the newspapers of India.

4. Theoretical Framework

'Towards A Taxonomy of Data Journalism-2017' authored by Charalampos Bratsas and Andreas Veglis focused on proposing a taxonomy of data journalism projects that can help the future journalists to choose the appropriate type of data project which will be suitable for their needs. In the research paper 'Visualization and Interactivity in Data Journalism Projects' authored by Christina Karypidou, Charalampos Bratsas, and Andreas Veglis, the researchers took into account the proposed taxonomy model of data journalism to determine the visualization effect and level of interactivity among the data journalism projects. Authors analyzed the websites of The Guardian and The New York Times - categorized as the elite newsrooms of the world, The British Broadcasting Corporation, and The Cable News Network-both of them are two most famous networks and the two largest news agencies namely, Associated Press and Reuters.

Figure 1: Taxonomy Model of Data Journalism (2017)
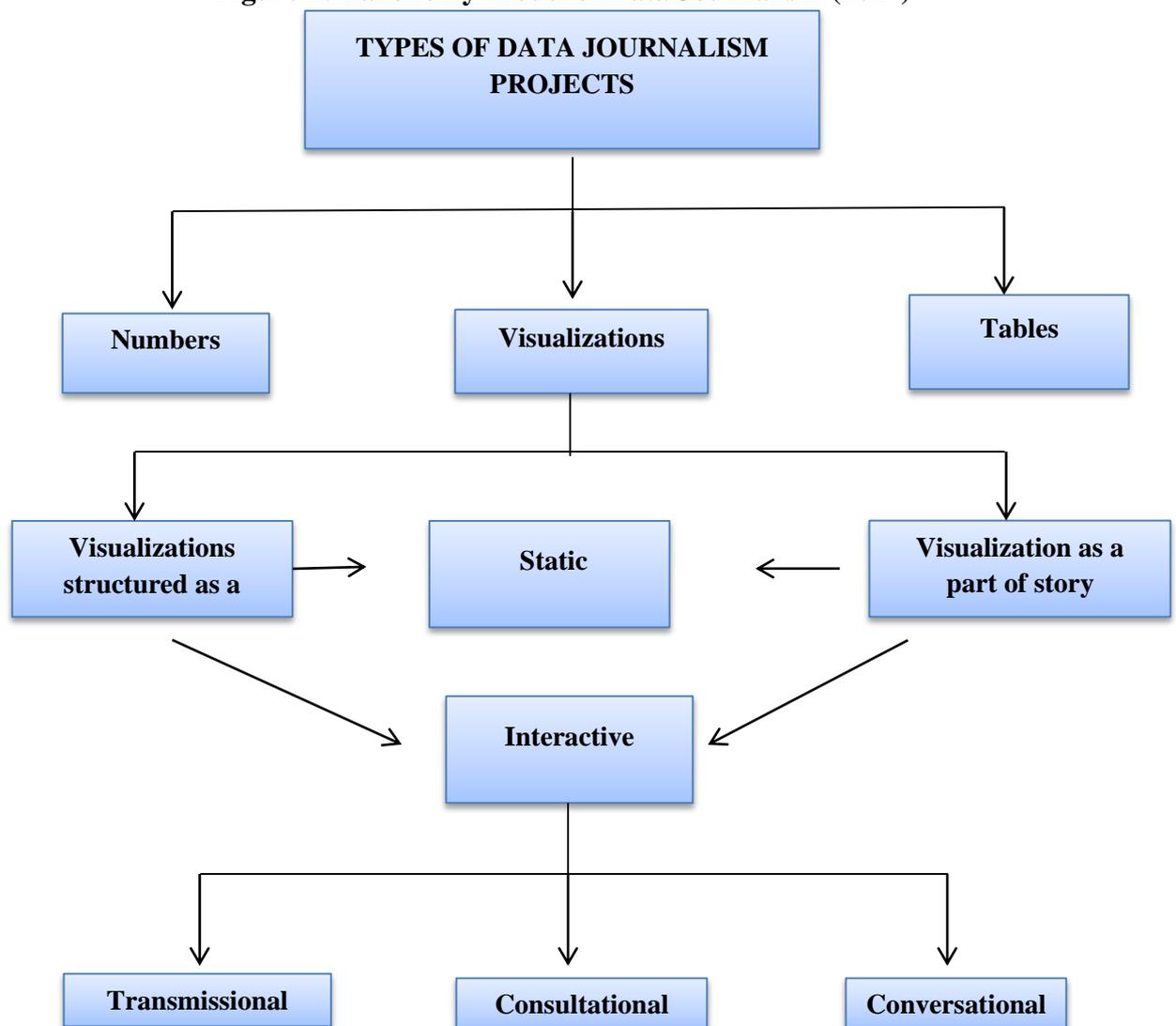
Unlike the Process Model of American Press Institute, Processing Model of Chicago Tribune (2013) and Workflow Model of Data Journalism (2017), the Taxonomy Model of Data Journalism (2017) cannot be fully implemented in this research due to the parameter of interactivity and interactivity level in the model, which is exclusively meant for the online platforms. Newspaper as a medium of communication is static, thus the Taxonomy Model of Data Journalism cannot be followed completely for the content analysis study of the newspaper.

Therefore, the researcher has made certain modifications in the parameters of the content analysis to support the proposed model of data-driven reporting in newspapers

1. Based on the classification of purpose and relationship
2. Based on the classification of data presentation, type, and structure

5. Methodology

The methodology used in this study by the researcher was a content analysis based on the quantitative approach. The researcher made use of unstructured sampling for the content analysis study which included an analysis of 3 national English daily newspapers of India namely, Times of India, Hindustan Times, and The Hindu. The sole objective of this research was to analyze the coverage of data-driven articles in the newspapers based on the Taxonomy Model of Data Journalism 2017 and to see if further modifications can be done to that model based on the practice of data journalism in the newspapers. According to the results of Indian Readership Survey-2019, the readership of English newspapers has increased from 2.7% to 2.9% in urban areas, whereas in the urban areas the readership of regional and Hindi newspapers remains stagnant. The researcher does not deny the fact that in India 70% of the population is situated in rural India and has maximum readership in Hindi dailies. However the study had to be completed within a given time frame, thus this study can further broader its unit of analysis if there are no time and resource limitations.

The duration of the study was from 5th January 2020 to 31st January 2020 taking into consideration newspapers of consecutive 10 days for each of the English dailies. The sampling plan of the study is given below-

Table 1: Sampling Plan Table

SNo	Newspaper Name	Dates
1	The Hindu	5th January-15th January 2020
2	Times of India	15th January -24th January 2020
3	Hindustan Times	20th January-31st January 2020

Source: Primary Data

The researcher on purpose chose to go forward with unstructured sampling for the content analysis study to avoid biases in the findings which might have occurred as in January 2020 main events of national importance were taking place like Delhi Elections, Republic Day, and Corona Virus Outbreak in the world. Therefore, to avoid any chances of biases in the representation of data-driven coverage in the newspapers, the researcher used an unstructured sampling plan for content analysis.

The coding categories for this content analysis were divided into 3 broad categories of the news articles. Given below is the codebook that was used for this study.

Once the codebook for the research was prepared the first set of code sheet which covered the analysis of data-driven news articles in The Hindu newspaper was sent on pretesting. This step was really important to determine whether the study was going in the right direction. It further helped in stipulating the findings of the research which will be obtained after analyzing the data-driven news articles of all the three newspapers.

The data for this study was collected from the 3 main national daily newspapers of India. In total 240 news articles were analyzed, the basis of analysis was The Taxonomy Model of Data Journalism-2017 suggested by Andreas Veglis and Charalampos Bratsas in their research paper titled 'Towards the Taxonomy Model of Data Journalism-2017'. The major software that was used for data analysis was Microsoft Excel and Statistical Package for the Social Sciences.

6. Findings of the Study

6.1 Findings based on the cross tabular data analysis and interpretation:

a.   Visibility of data-driven news stories in various newspapers: Through the content analysis methodology used in this research, it was concluded that Times of India covered

highest number of data-driven news stories with 40.4%, followed by The Hindu with 37.1% and the least was found in the Hindustan Times with 22.5% of data-driven coverage.

Figure 2: The graph represents the visibility of data-driven news stories in various newspapers
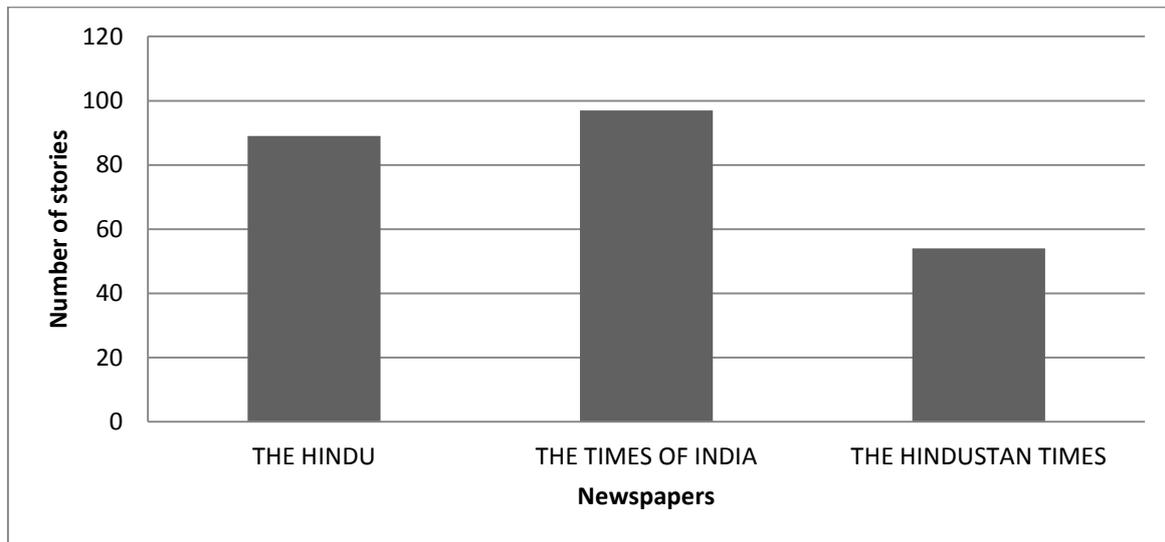
Source: Primary Data

b.   Innovative display of data-driven news stories in various newspapers: In the content analysis it was found that Times of India presented most of its data-driven news stories in an innovative form with 77% out of 240 followed by Hindustan Times with 6%. However, The Hindu had no such practice.

c.   The special section included for data-driven news stories in various newspapers: It was found in the analysis, that Hindustan Times had the majority of data-driven stories with 15% out of 54 data-driven news stories in a special section of the newspaper namely, 'The Number Theory' followed by the Hindu newspaper, with 7% out of 89 data-driven stories in a special section for the data-driven news stories, namely 'The Data Point'.

d.   Report formats are the main carriers of data-driven news stories: In all three national dallies, news report format was the main carriers of data-driven news stories with 92.13% in The Hindu. 96.90% in Times of India and 54 out 54 data-driven stories of Hindustan Times were covered in news report format.

e.   The dominance of the beat in covering data-driven news stories: For the Hindu, the majority of data-driven stories were covered in sports beat with 11.23% out of 89 stories. Times of India covered the majority of data-driven news stories in the economic beat with 32.98%

out of 97 data-driven stories. In the case of Hindustan Times the majority of data-driven news stories were covered in sports and national beat with 31.48% out of 54.

Figure 3: The graph represents the dominance of the beat in covering data-driven news stories
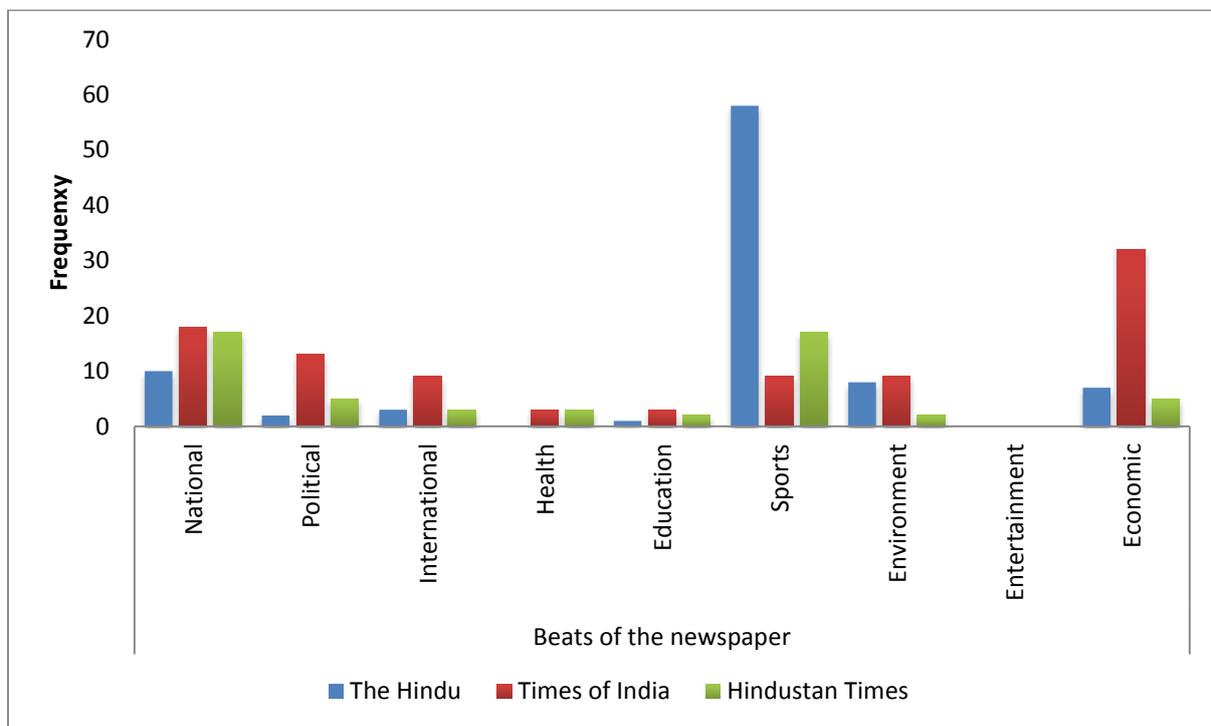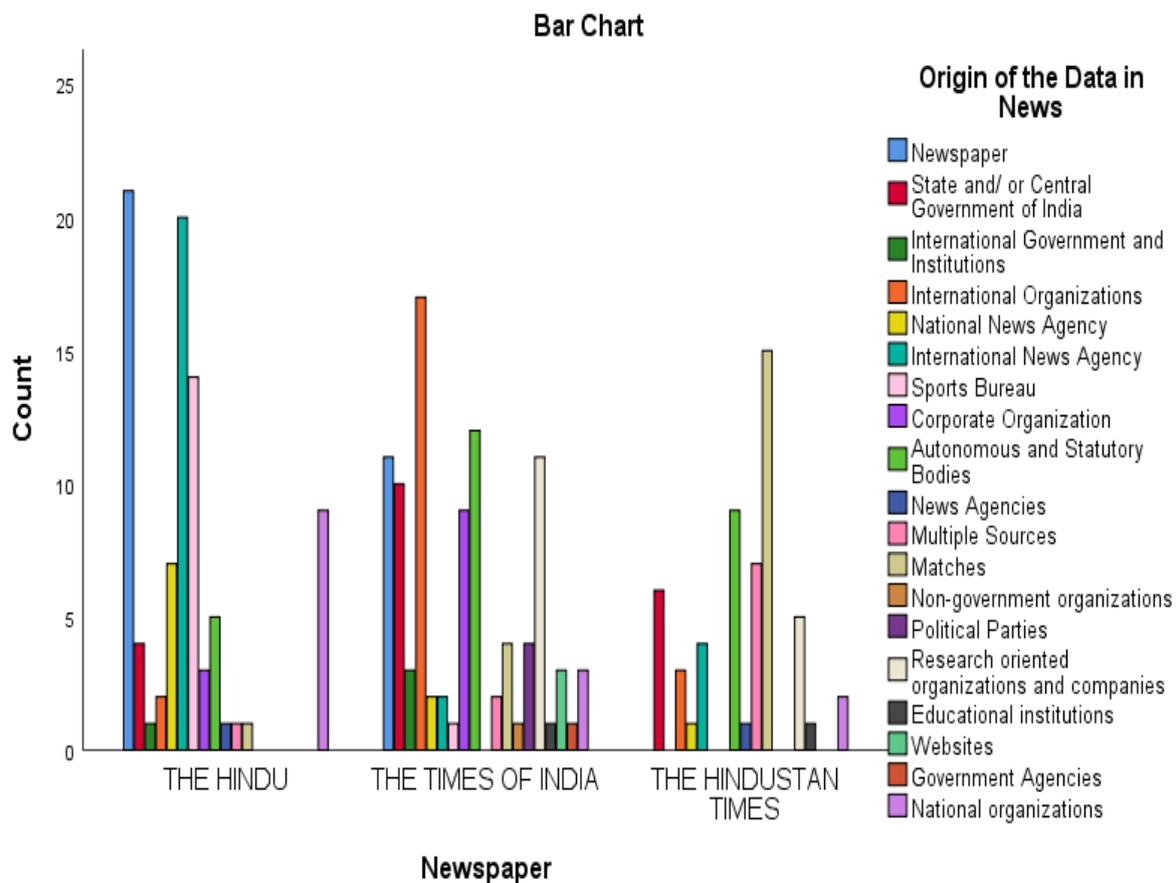

Source: Primary Data

f.   Prominent news sources for covering data-driven news stories in 3 national dailies: For the Hindu, most of the data-driven news stories were covered by news agencies with 33.70% out of 89 news stories and the least was reported by the unnamed reporters and/or correspondents with 30.33% out of 89 data-driven stories. For Times of India, the highest numbers of data-driven news stories are reported by their named reporters with 51.4% out of 97 data-driven stories and correspondents with 34.02%; however Times of India makes the least use of data-driven news stories directly covered by the news agencies with 14.43% out of 97 data-driven news stories. In the case of Hindustan Times named reporters and/or correspondents with 66.66% out of 54 data-driven news stories covered the majority of data-driven news stories and least are covered by their unnamed reporters and/or correspondents with 12.96%

g.   Importance of news agencies for coverage of data-driven news stories in major 3 national dailies: As per the content analysis study, The Hindu is majorly dependent on the news agencies for their coverage of data-driven news stories. However, in the case of the Times of India and Hindustan Times, it is only 14.43% and 20.37% of data-driven stories that are directly taken by the news agencies respectively.

h.   Major origin of data for all three national dailies: For the Hindu and Times of India their major origin of data was from their newspaper organization with and reporters and/or correspondents play a major role in gathering, analyzing, and compilation of data for news stories. However, this was not found in Hindustan Times. Hindustan Times mostly relies on the data released by the State and the Central Government of India with 11.11% out of 54 analyzed news stories. The second-largest data source for The Hindu and Times of India is the international news agencies and organizations respectively.

Figure 4: The graph represents the origin of data for all three national dailies

Source: Primary Data

i.   Variation in the national dailies based on the purpose of data: The purpose of data varied from one newspaper to another. The data-driven news stories of The Hindu majorly focused on just mentioning facts with 71.9% out of 89 news stories, whereas data-driven news stories of Times of India were more analytical with 67.01% out of 97 news stories. In the case of the Hindustan Times, analytical data-driven news stories were 10% more of data-driven news stories that just mentioned fact.

Figure 5: The graph represents variation in the national dailies based on the purpose of data
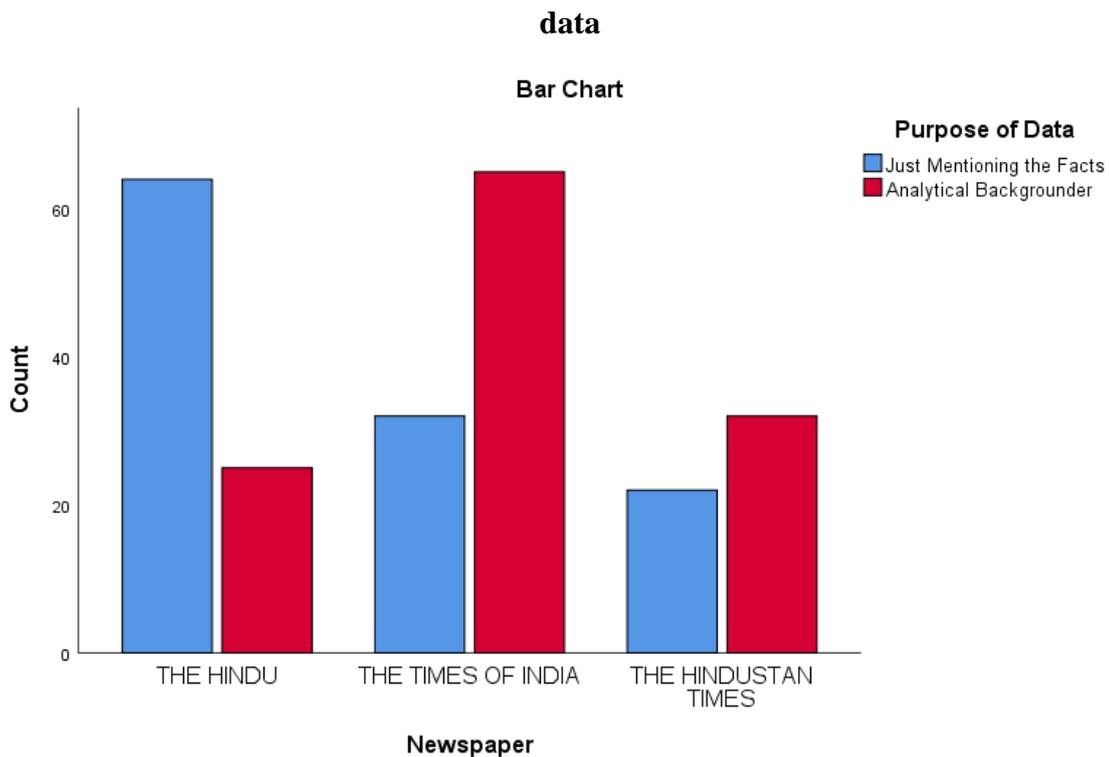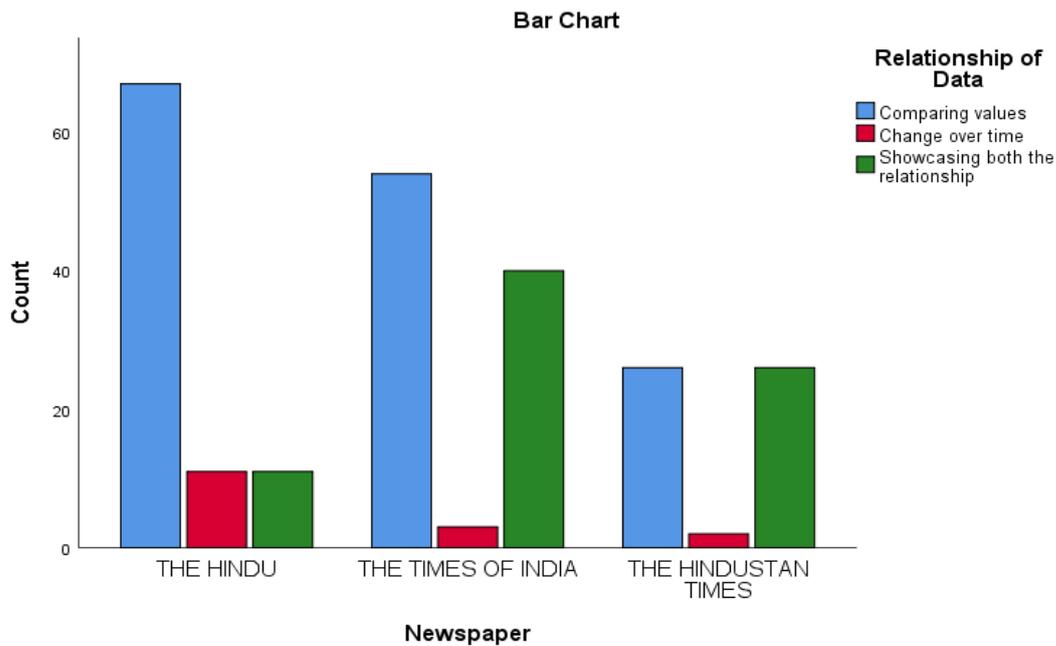

Source: Primary Data

j.   Variation in the national dallies based on the relationship of data: In the Hindu majority of data presented relationship of comparison in the data-driven news stories with 71.9%. Data showing change over time relationship and combined relationship were equal in the case of The Hindu with 12.35% out of 89 news stories. For Times of India, the data represented the relationship of comparison majorly through the news stories with 41.22% out 97 news stories. Followed by presenting a combined relationship, however, data representing change overtime were least to be found in the news stories of Times of India. Hindustan Times through its news reports showcased data establishing a relationship of comparison and combination of both the

relationships were at an equal pace with 26 news stories out of 54. However, Hindustan Times established the least amount of data showing change over time relationships with 6.6%.
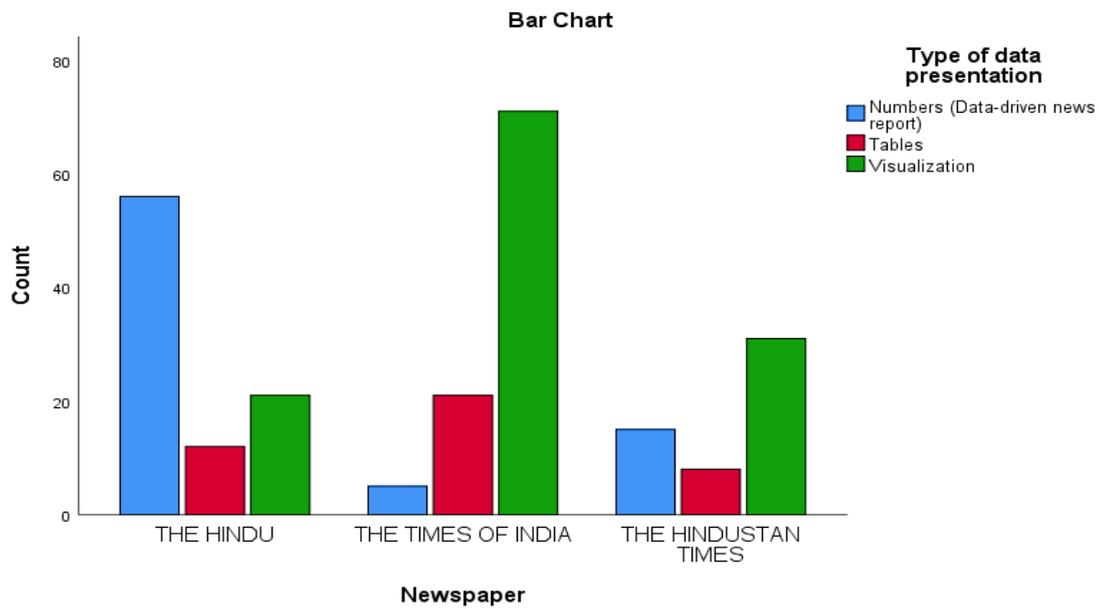
Figure 6: The graph represents variation in the national dallies based on the relationship of data

Source: Primary Data

k.   Variation in the national dailies based on the type of data presentation: As per the content analysis results, it was concluded that The Hindu majorly presented its data in the form of news reports, followed by visualizations and tables. However, both Times of India and the Hindustan Times presented their data through the mode of visualizations. For Times of India second highest presentation was through tables followed by news reports. In the case of the Hindustan Times, the second-highest representation was of news reports followed by the tables.
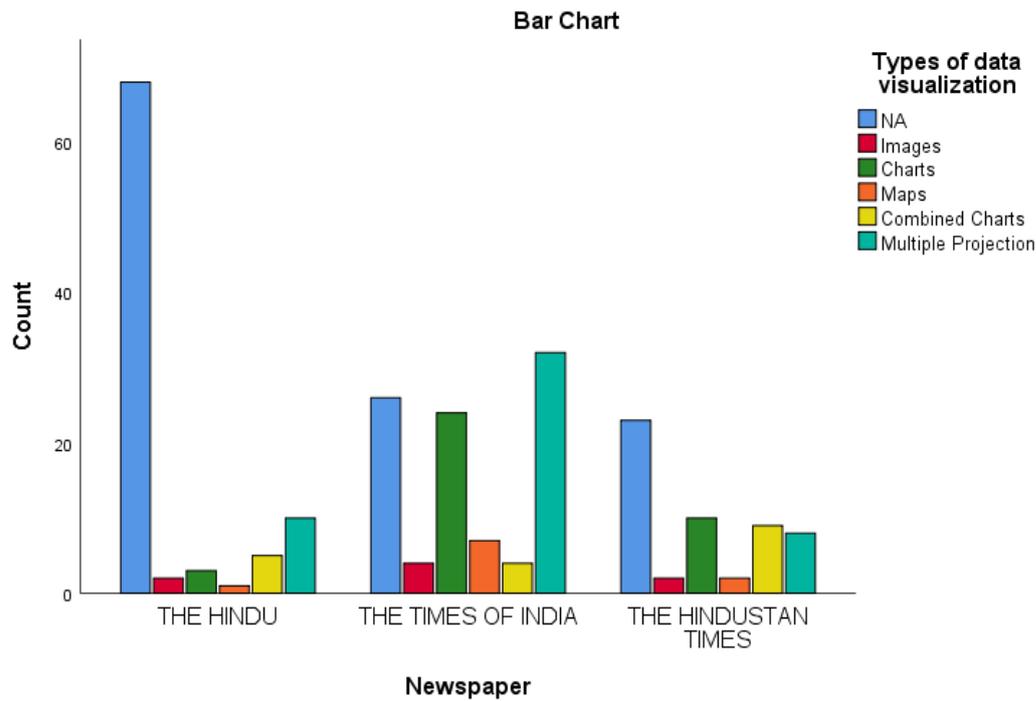
Figure 7: The graph represents variation in the national dailies based on the type of data presentation

Source: Primary Data

l.   Variation in the national dailies based on the type of data visualization: Through the content analysis methodology it was figured out that The Hindu visualizations were in majority represented through multiple projections and combined charts. However, the least was presented through images. Times of India represented visualizations through multiple projections mainly followed by analytical charts. The least number of visualizations were presented through images and combined charts. In the case of the Hindustan Times majority of the data, visualizations were presented in the form of charts, followed by combined charts and multiple projections. The least amount of data was represented through images and maps.

Figure 8: The graph represents variation in the national dailies based on the type of data visualization

Bar Chart

Source: Primary Data

m. Variations in the national dallies based on the structure of visualization: As per the results of the content analysis all 3 national dailies covered data-driven stories wherein major data was structured as a story other than data visualizations which remained just a part of the story.

Figure 9: The graph represents in variations in the national dallies based on the structure of the visualization
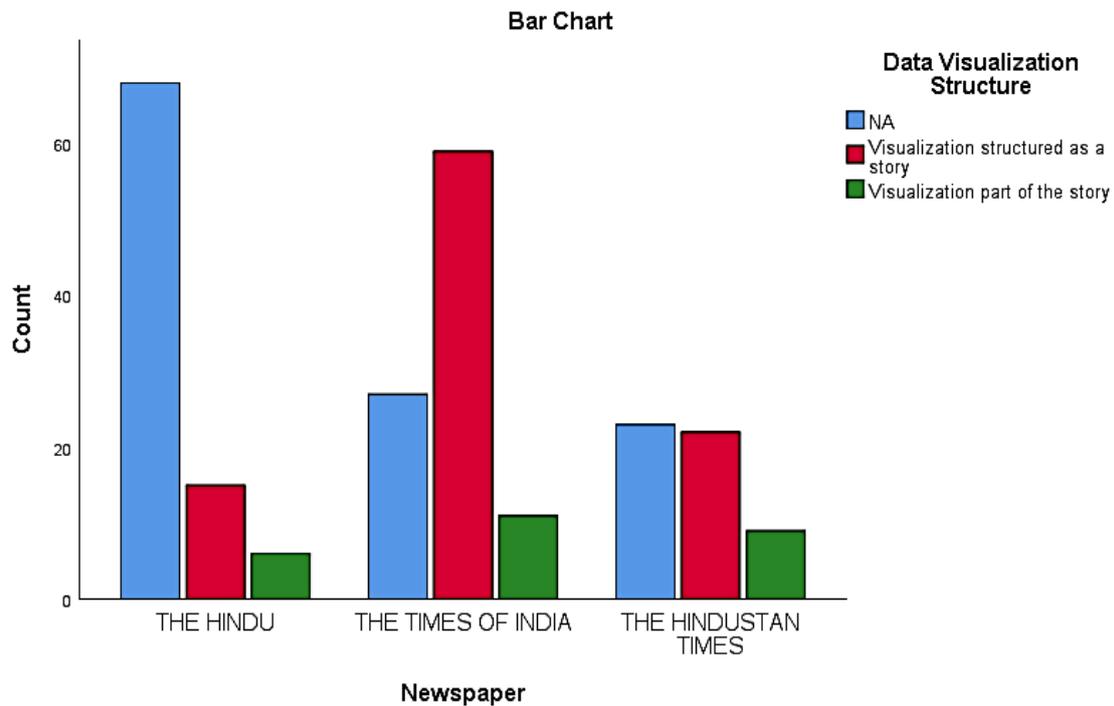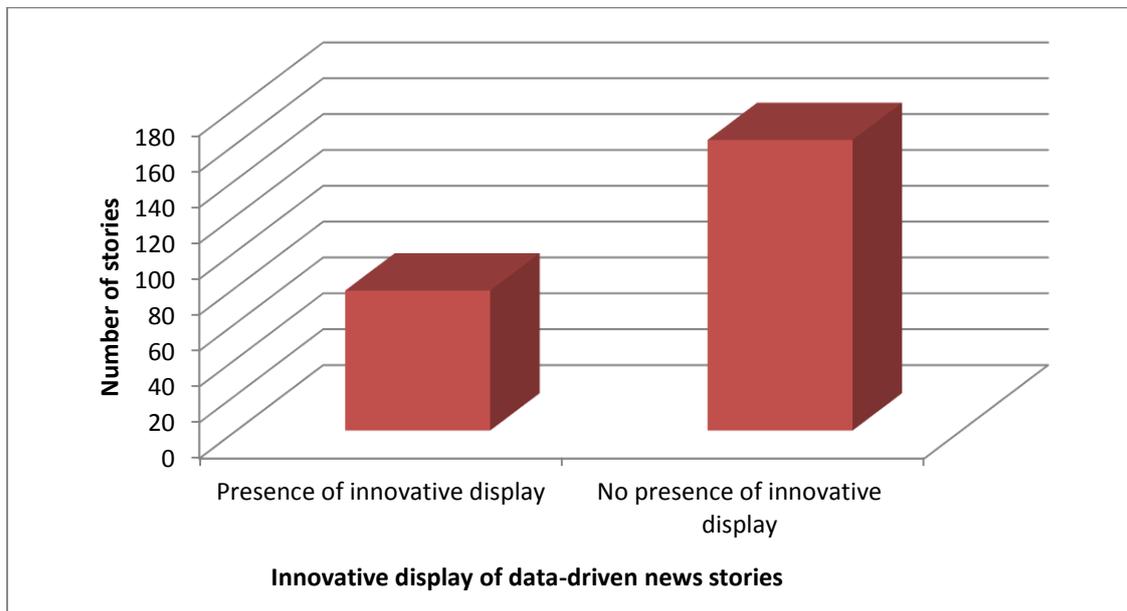

Source: Primary Data

6.2 Findings based on collective data analysis and interpretation:

a.   Primary news sources in data-driven news stories: As per the results of content analysis of 240 data-driven news stories it can be concluded that major news sources for data-driven stories are the named reporter and/or correspondents, followed by the unnamed reporters and/or correspondents and the least were the news agencies 22.9% of the data-driven stories out of 240 were taken from the news agencies. However, for 1.7% of the news stories, no news sources were mentioned.

b.   Presence of innovative display in data-driven news stories: For the majority of the data-driven news stories there was no innovative display in all the three newspapers.

Figure 10: The graph depicts the presence of innovative display in data-driven news stories

Source: Primary Data

c.   Presence of special section for data-driven news stories: For a majority of data-driven news stories there was no inclusion of special section for data-driven news stories in all the three newspapers.

Figure 11: The graph depicts the presence of a special section for data-driven news stories
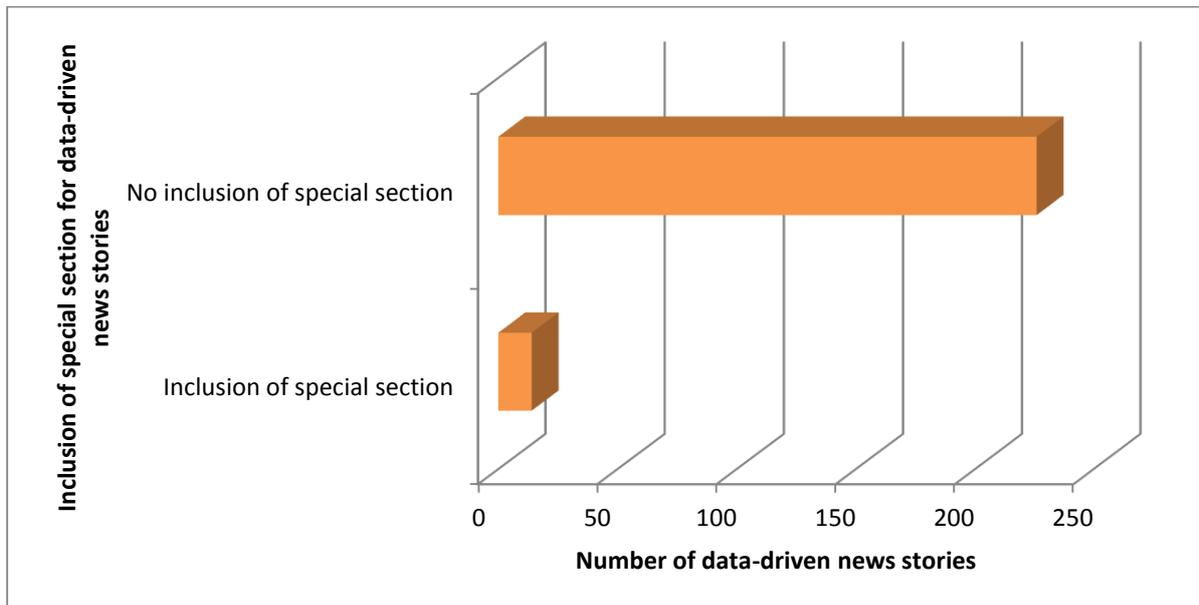
Source: Primary Data

d.   Format dominantly used to carry out data-driven news stories: According to the study conducted it was concluded that the majority of the news stories are covered in the form of the news report and very minimal data-driven stories are categorized under the editorial and opinion piece section of newspapers.

e.   Dominant beat in carrying out data-driven news stories: As per the analysis majority of the data-driven news stories were covered in sports beat collectively. Further in the majority are the data-driven news stories falling under the economical and national beat of the newspapers.

Delhi Metropolitan Education
Affiliated to GGSIP University, New Delhi & Approved by Bar Council of India

Figure 12: The graph depicts the dominant beat in carrying out data-driven news stories
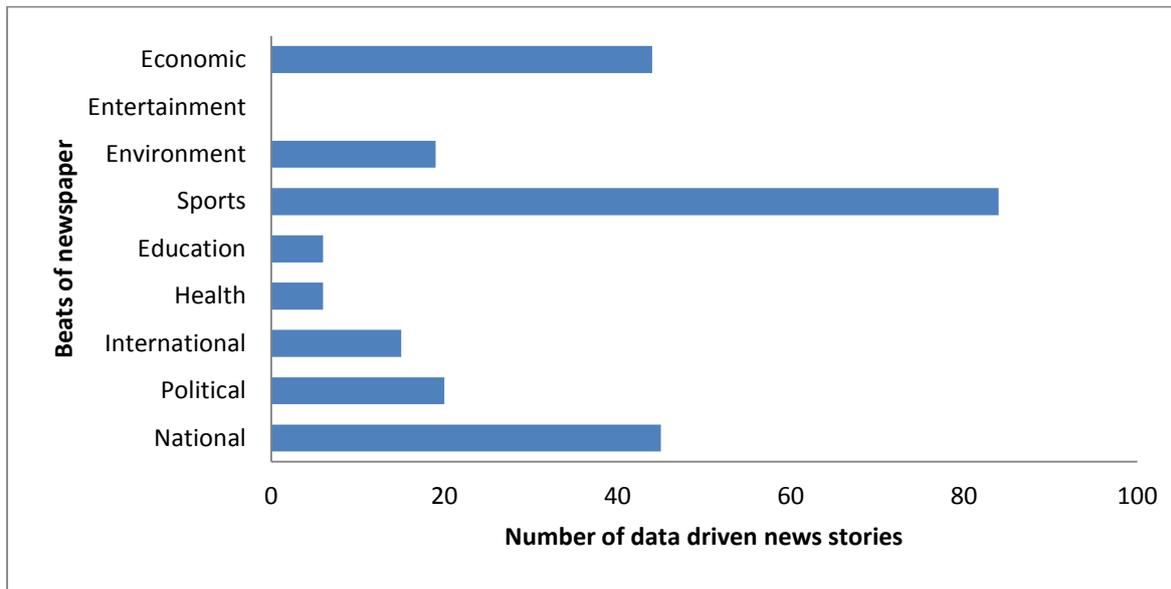
Source: Primary Data

f.   Primary data source in the data-driven news stories: Through the content analysis methodology it was figured out the main source for data are the newspaper organizations itself. Further in the majority is data generated from international news agencies and autonomous bodies. However, newspapers used the least of the data which was obtained from websites.

Figure 13: The graph depicts the primary data source in the data-driven news stories
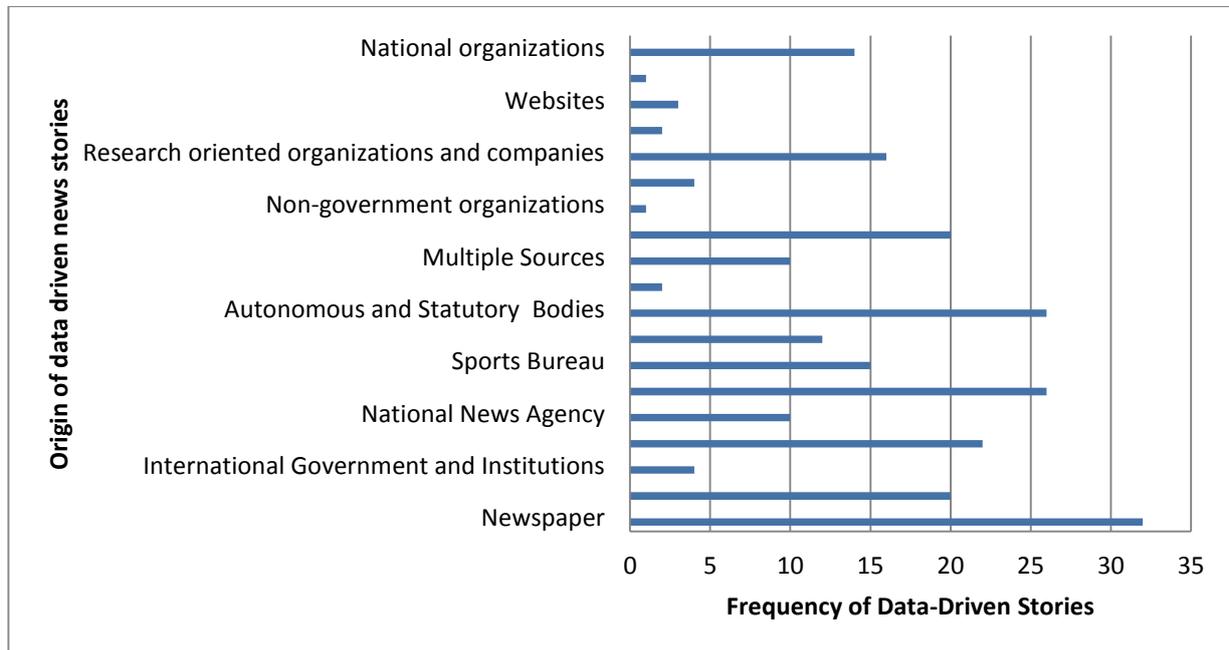
Source: Primary Data

g.   The dominance of data-driven stories based on purpose: Through this study, it was figured that newspapers focus more on analytical based data-driven stories as compared to those stories which just mention facts.

Figure 14: The graph depicts the dominance of data-driven stories based on the purpose
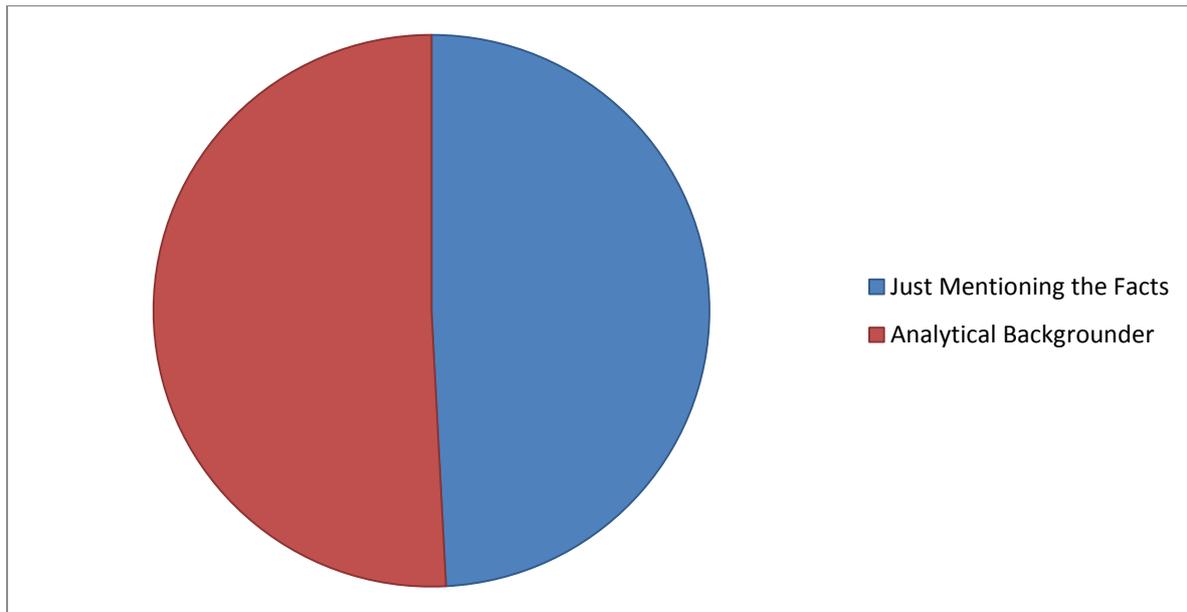
Source: Primary Data

h.   The dominance of data-driven stories based on relationship: It was concluded that the majority of the data in the news stories showed the relationship of comparison, followed by stating both the relationships of comparison and change over time. However, data showing change over time relationships were less in the 240 data-driven news stories which were analyzed.
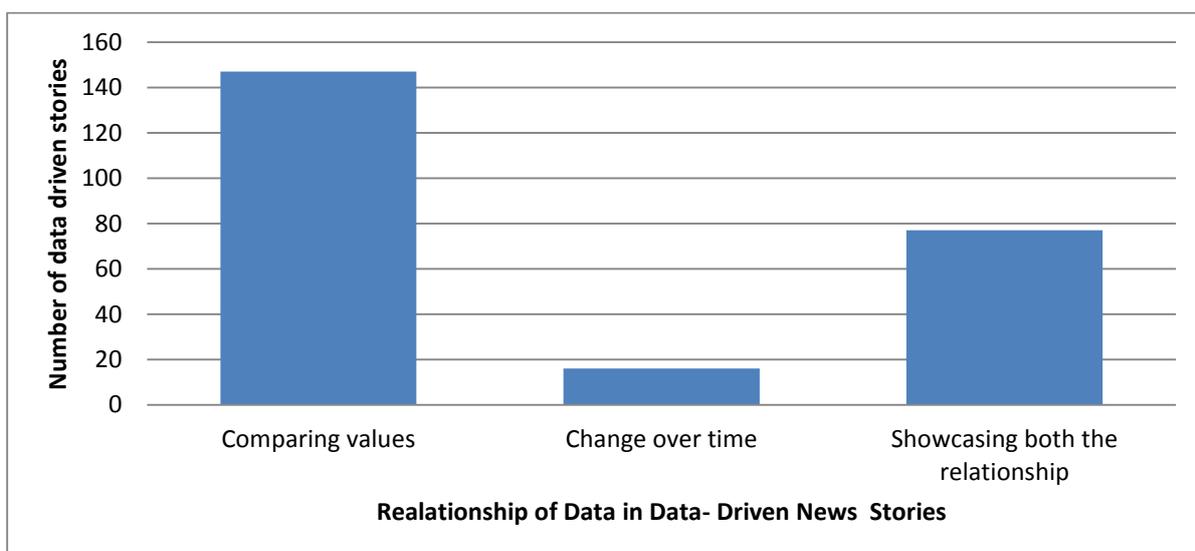
Figure 15: The graph depicts the dominance of data-driven stories based on the relationship

Source: Primary Data

i.   The prominence of presentation of data in data-driven news stories: In the majority of the news stories data was presented through the mode of visualization. The second highest was the presentation of data in the form numbers mentioned in the news report and least were in the case data presented in the form of tables.

Figure 16: The graph depicts prominence of presentation of data in data-driven news stories
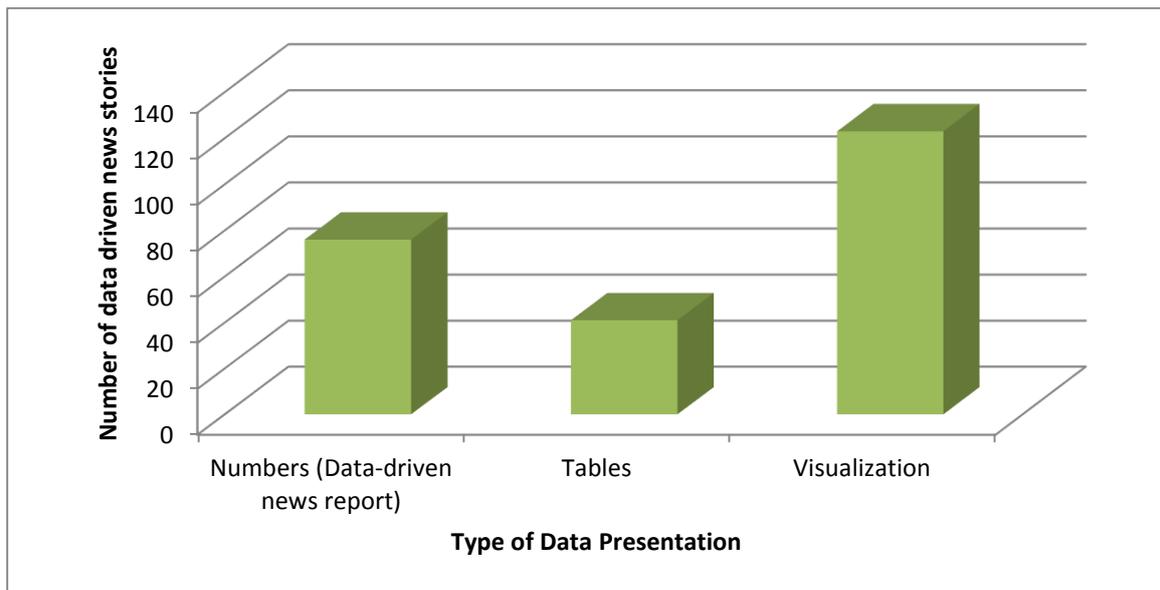
Source: Primary Data

j.   The dominance of type in the data visualizations within the data-driven stories: As per the results of the content analysis, in the majority of the news articles data was published in the form of multiple projections, followed by charts and combined charts. The data was least presented in the form of images and maps. However, for 48.8% of data-driven news stories, this criteria wasn't applicable as data was presented either in the form of news reports or tables.

Figure 17: The graph depicts the dominance of type in the data visualizations within the data-driven stories
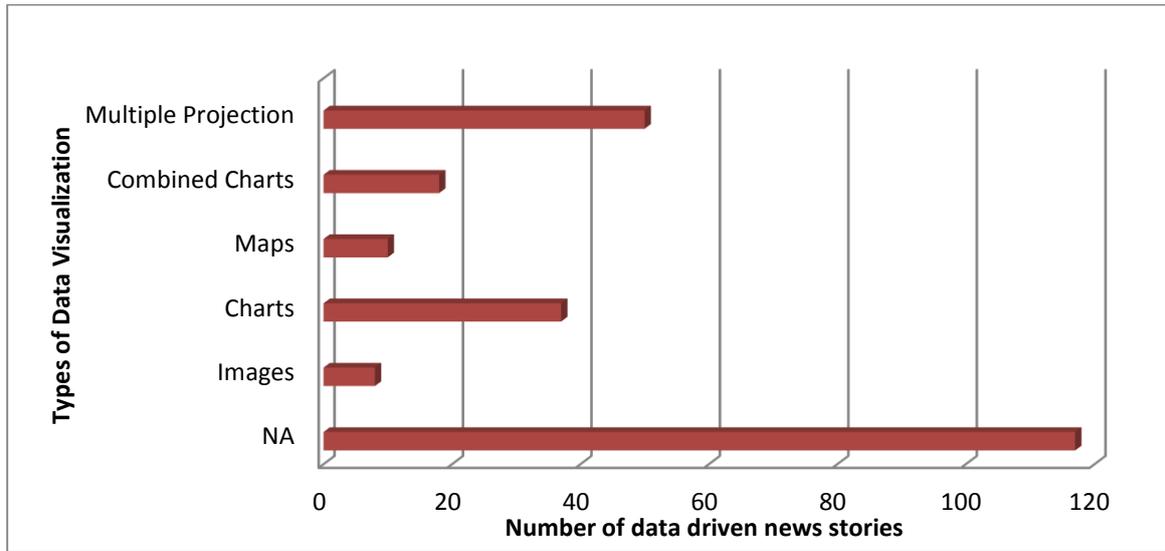
Source: Primary Data

k.   The dominance of the visualization structure within the data-driven stories: According to the analysis, the majority of the data visualizations were structured as a story as compared to data visualizations which just acted as a part of the story. But for 48.8% of data-driven news stories, this criteria wasn't applicable as data was presented either in the form of news reports or table.

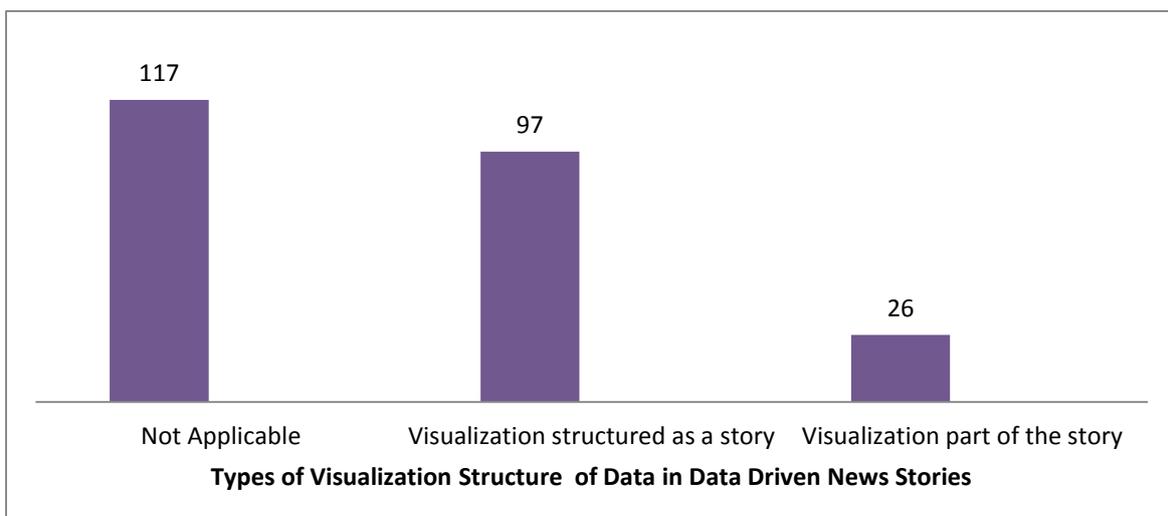**Figure 18: The graph depicts the dominance of the visualization structure within the data-driven stories**

Source: Primary Data

7. Models Suggested for Data-Driven News Reporting in Newspapers

In this model, the data-driven news stories of newspapers were initially classified based on purpose which included data-driven news stories with the sole purpose of just mentioning facts and remaining which depicted analytical aspects with the help of data. Further, the data-driven news stories can be classified based on the relationship which included parameters like data stating comparing values, change over time or either stating both the relationships.

Figure 19: Proposed model of data-driven reporting in newspapers based on purpose and relationship
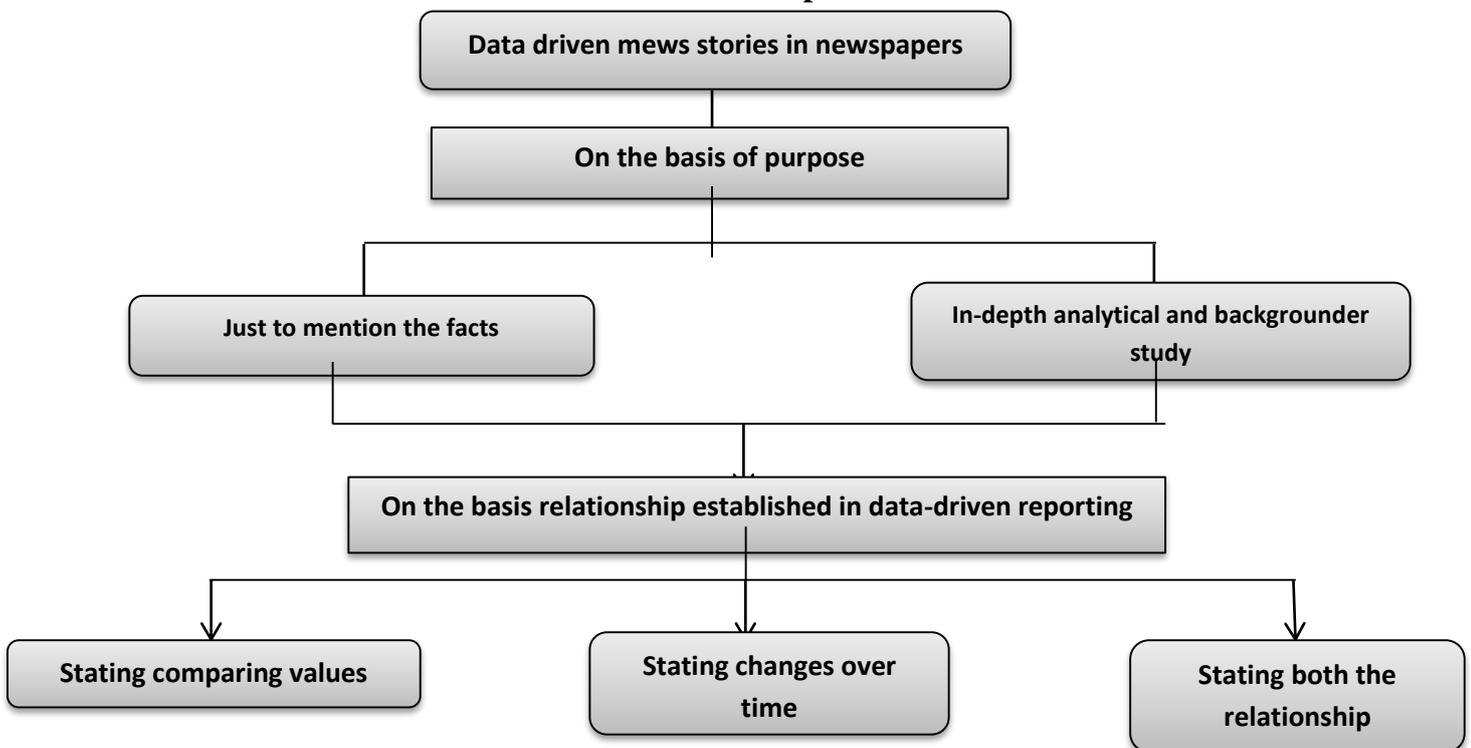
Source: Primary Data

In the second model, the data-driven news stories were classified based on data presented in the data-driven news stories. The classification included parameters like data represented through news reports, tables, and visualizations. The data visualizations were further classified into types of data visualizations which included images, charts, maps, combined charts, and multiple projections or visualizations. The data visualizations are further classified based on structure wherein visualizations are either structured as a story or it's a part of the story.

Figure 20: Proposed model of data-driven reporting in newspapers based on the classification of data presentation, type, and structure

```
            ┌─────────────────────────────────────────────────┐
            │  Presentation of data driven coverage in newspaper │
            └─────────────────────────────────────────────────┘

┌──────────────────────┐   ┌──────────────────────┐   ┌──────────────────────┐
│ Numbers in data driven│   │ Visualization in data │   │  Tables in data driven│
│  coverage of          │   │  driven coverage      │   │  coverage of newspaper│
│  newspaper            │   │  of newspaper         │   │                       │
└──────────────────────┘   └──────────────────────┘   └──────────────────────┘

            ┌─────────────────────────────────────────────────┐
            │ Classification of data visualization in newspaper │
            │           on the basis of type                    │
            └─────────────────────────────────────────────────┘

 ┌────────┐   ┌────────┐   ┌────────┐   ┌────────────────┐   ┌──────────────┐
 │ Images │   │ Charts │   │  Maps  │   │ Combined Charts │   │   Multiple   │
 └────────┘   └────────┘   └────────┘   └────────────────┘   │ Visualizations│
                                                             └──────────────┘

            ┌─────────────────────────────────────────────────┐
            │ Classification of data visualization in newspaper │
            │       on t-he basis of structure                  │
            └─────────────────────────────────────────────────┘

 ┌──────────────────────────────┐          ┌──────────────────────────────┐
 │ Visualization structured as a │          │ Visualization as part of a    │
 │           story               │          │           story               │
 └──────────────────────────────┘          └──────────────────────────────┘
```

8. Conclusion of the Study

The main aim of this research was to analyze the coverage of data journalism in three main daily papers of India based on the taxonomy model of data journalism-2017. The main aim of this paper was to investigate if the model was suitable for data-driven coverage in the newspapers and if not were there any modifications that can be proposed as an extended model of this taxonomy based on the practice of data journalism in the newspapers.

The extended model of data journalism was proposed to help print journalists to follow the more structuralized model of data journalism which will cater to their needs and will open a new chapter in educational strategies of data journalism. The proposed model proved that data journalism as a term is not limited to online media platforms, but can be implemented within the print media and can raise similar challenges to print journalists. Hence data journalism and its educational strategies should not be limited to online media platforms, as in the age of transparency and authenticity newspapers are depending heavily on data to maintain the genuineness of the medium.

The positive aspect of this research study was that the researcher took into consideration the purpose of the data, the relationship established through data, and lastly the type of data visualizations utilized by various newspapers. For the first time practice of data journalism in the newspapers was put into focus rather than limiting its scope to online platforms. Hence all the parameters helped the researcher propose models of data-driven news stories in the newspapers.

The basic limitation of this study can be that the researcher could have increased the sample of the study for more scope of generalization. However due to the limitation of time and resources the study was confined to an unstructured sample of ten consecutive newspapers of each national dailies.

The future scope of the study for other researchers can be to identify the further classification of data presentation with news reports and tables. The new classification can further modify the scope of the model and its practice can even be tested in the Hindi newspapers which has the highest readership in India. This exercise will only provide more authenticity to the models or can even bring up new findings.

References

Bradshaw, Samantha & Philip N., Howard (2017). *Troops, Trolls and Troublemakers: A Global Inventory of Organized Social Media Manipulation*. Accessed from https://comprop.oii.ox.ac.uk/ research/troops-trolls-and-trouble-makers-a-global-inventory-of-organized-social-media-manipulation/

Campbell-Smith, Ualan & Bradshaw, Samantha (2019). *Global Cyber Troops Country Profile: India*. Oxford Internet Institute, University of Oxford https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/ 2019/05/India-Profile.pdf

Carr, Austin et.al (2019). *Silicon Valley is listening to your most intimate moments*. Accessed from https://www.bloomberg.com/news/features/2019-12-11/silicon-valley-got-millions-to-let-siri-and-alexa-listen-in

Collins, Keith & Dance JX ,Gabriel (2018). *How Researchers Learned to Use Facebook 'Likes' to Sway Your Thinking*. https://www.nytimes.com/2018/03/20/technology/facebook-cambridge-behavior-model.html

Confessore, Nicholas (2018). *Cambridge Analytica and Facebook: The Scandal and the Fallout So Far*. Accessed from https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html

Dreyfuss, Emily (2018). *Google Tracks You Even If Location History's Off. Here's How to Stop It*. Accessed from https://www.wired.com/story/google-location-tracking-turn-off/

Flaxman, Seth et.al (2016). *Filter Bubbles, Echo, Chambers, and Online News Consumption.*

Harari, Y. (2017) *Homo Deus - A Brief History of Tomorrow*, (rev. ed) London: Vintage (Penguin Random House)

Hogan,Mél & Shepherd, Tamara (2015). *Information Ownership and Materiality in an Age of Big Data Surveillance*. Journal of Information Policy , 2015, Vol. 5 (2015), pp. 6-31. Published by: Penn State University Press. Stable URL: https://www.jstor.org/stable/10.5325/jinfopoli. 5.2015.0006

Ingraham, Christopher (2018). *An insurance company wants you to hand over your Fitbit data so it can make more money. Should you?*Accessed from https://www.washingtonpost.com/business /2018/09/25/ an-insurance-company-wants-you-hand-over-your-fitbit-data-so-they-can-make-more-money-should-you/

Lapowsky, Issie (2018). *The Man Who Saw the Dangers of Cambridge Analytica Years Ago*. Accessed from https://www.wired.com/story/the-man-who-saw-the-dangers-of-cambridge-analytica/

Leetaru, Kalev (2018). *Facebook's Automated Ad Labels Plus Facial Recognition: The Real-Life Minority Report?* Accessed from https://www.forbes.com/sites/kalevleetaru/2018/07/18/facebooks-automated-ad-labels-plus-facial-recognition-the-real-life-minority-report/#2ab799176808

Leetaru, Kalev (2019). *Global Mass Surveillance And How Facebook's Private Army Is Militarizing Our Data*. Accessed from https://www.forbes.com/sites/kalevleetaru/2019/03 /11/global-mass-surveillance-and-how-facebooks-private-army-is-militarizing-our-data/#31a25cd81786

Levitin, Daniel J. (2016). *Weaponized Lies: How to think critically in the Post-Truth Era.* Dutton, New York

Meredith, Sam (2018) *Here's everything you need to know about the Cambridge Analytica scandal.* Accessed from l https://www.cnbc.com/2018/03/21/facebook-cambridge-analytica-scandal-everything-you-need-to-know.html

Michal Konsinski et al. (2013). *Private traits and attributes are predictable from digital records of human behavior.* https://www.pnas.org/content/pnas/110/15/5802.full.pdf

Orwell, George (Reprint 2018). *1984.* Published by Prakash Books India Pvt. Ltd.

Public Opinion Quarterly, Vol. 80, Special Issue, 2016, pp. 298–320

Singh Shankar, Shivam (2019). *A former BJP data analyst reveals how the party's WhatsApp groups work.* Accessed from https://qz.com/india/1553765/bjps-whatsapp-ops-is-what-cambridge-analytica-can-only-dream-of/?utm_source=facebook&utm_medium=qz-organic

Stephens –Davidowitz, Seth (2017). *Everybody lies: What the internet can tell us about who we really are.* Bloomsbury Publishing, London, UK

Vaidhyanathan, S. (2018) Antisocial Media - How Facebook Disconnects Us and Undermines Democracy, New Delhi: Oxford University Press

Withnall, Adam (2016). *Uber knows when your Phone is running out of Battery.* Accessed from https://www.independent.co.uk/life-style/gadgets-and-tech/news/uber-knows-when-your-phone-is-about-to-run-out-of-battery-a7042416.html