Impact Factor (SJIF): 3.632

International Journal of Advance Research in Engineering, Science & Technology

e-ISSN: 2393-9877, p-ISSN: 2394-2444

Volume 3, Issue 10, October 2016

Big Data Visualization Using Excel Features

Ruchita Jani[1]

P.G. Student, Department of Computer Engineering, Atmiya Institute of technology and science College, Rajkot, Gujarat, India[1]

ABSTRACT: Now a days, data is growing in every field with a way faster rate due to growth of social media. So, visualization is an important approach to helping big data get a complete view of data that can help decision making. There are many tools for visualizing data in graphical format. Among all, Excel is the most used tool by industry. Now a days it has been added with lots of new features so that it can work with big data as well. So, the most used tool for visualization, its challenges and improvements in in context of big data is introduced in this paper.

KEYWORDS: Big Data Visualization, Visualization Methods, Microsoft Excel, New Features, Limitations.

I. INTRODUCTION

In this technology trend, Data is growing way faster than before. People are sharing their information easily in every field on internet via social media. To process this huge amount data, there are lots of tools developed. Big data [1] Visualization is a representation of data in pictorial or graphical format. It enables us to evaluate things visually so that we can identify new patterns. Now, the main issue is visualizing massive data in order to understand and extract knowledge from stored data and give efficient decision making through its reports. In simple words, [1] the collection of information is no longer a problem but extracting of valuable knowledge from the collected or available information is the main thing.

Visualization methods are very important as it provides mental model of data to users. It makes huge and complex information intelligible. [1] The basic purpose of visualization is to create interactive visual representations of the huge information that increase capability of problem solving of human. User should be able to easily understand and interpret huge and complex set of information.

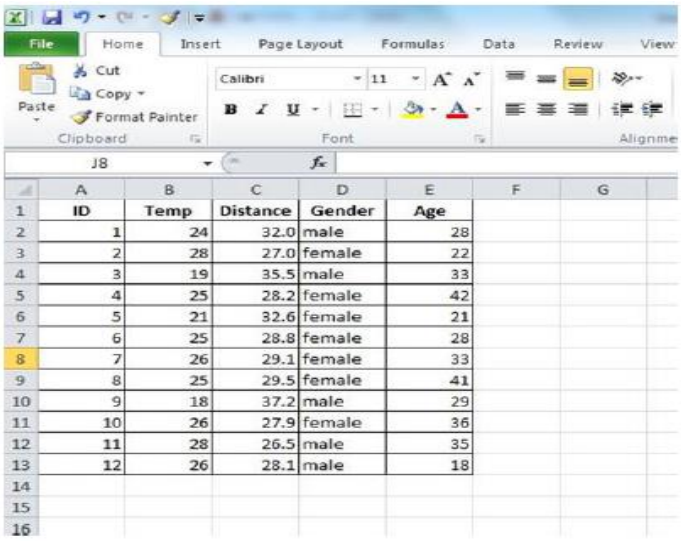
There are lots of tools for visualizing information. Visualization Approaches [2] are the techniques used to create tables, images, diagrams and other intuitive display ways to understand data. Big Data visualization is not that easy like traditional relative small data sets because of the complexity in big data. The extension of traditional visualization approaches are already emerged but far away from enough. Among all, here we are going to cover one tool that is Microsoft Excel, the most popularly used tool by IT industries and individuals as well. Microsoft Excel is very easy to use but it is not that smaller we think. It is having many features inside it. It has launched with effective and powerful new features to make better representation of massive data. Still it has some problems that leads to a challenge and need extension. The main objective of this paper is to provide basic knowledge about various methods and progress of visualization techniques. This also includes the future enhancements needed to improve current scenario.

II. VISUALIZATION METHODS

A number of visualization techniques have been developed in last few years due to need of representing huge amount of data and providing better decision making from these reports. All the tools mostly provide some basic kind of conventional techniques that make user easily understand the data in brief. So, we are going to discuss few known traditional techniques and give its example in order to get initial idea about features with Microsoft Excel tool.

1. Table

Table is a very basic, simple and easy to implement, easy to understand and interpret data representation technique. Table [1] is structured format including rows and columns. Row has different synonyms like tuple, vector etc. Columns are also referred as field, attribute etc. The role of table is key part in research and analysis. Example of excel table is as given in below figure.



	A	B	C	D	E	F	G
1	ID	Temp	Distance	Gender	Age		
2	1	24	32.0	male	28		
3	2	28	27.0	female	22		
4	3	19	35.5	male	33		
5	4	25	28.2	female	42		
6	5	21	32.6	female	21		
7	6	25	28.8	female	28		
8	7	26	29.1	female	33		
9	8	25	29.5	female	41		
10	9	18	37.2	male	29		
11	10	26	27.9	female	36		
12	11	28	26.5	male	35		
13	12	26	28.1	male	18		
14							
15							
16							

Fig 1: Showing simple table

2. Pie Chart

Pie chart is also called the circle graph. [1] Pie chart circle include number of sectors, each sector or portion describe a proportion in a whole quantity. It is mainly used to determine size of particular part of common data with other parts of data. Mostly sector is shown by percentage as shown in below figure.

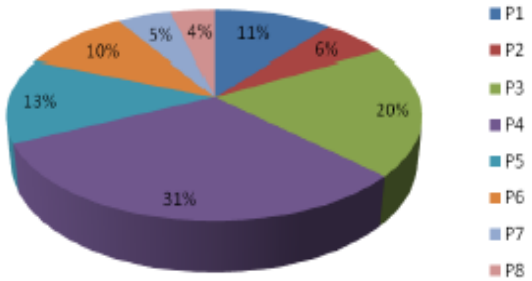


Fig 2: Standard pie chart

3. Bar Chart

Bar chart is one of the most commonly used method for representing data mostly for discrete data rather than common data. Bar chart is the diagram in which numerical values of variables are represented by the height or length of lines or rectangles of equal width as shown in below figure [3].

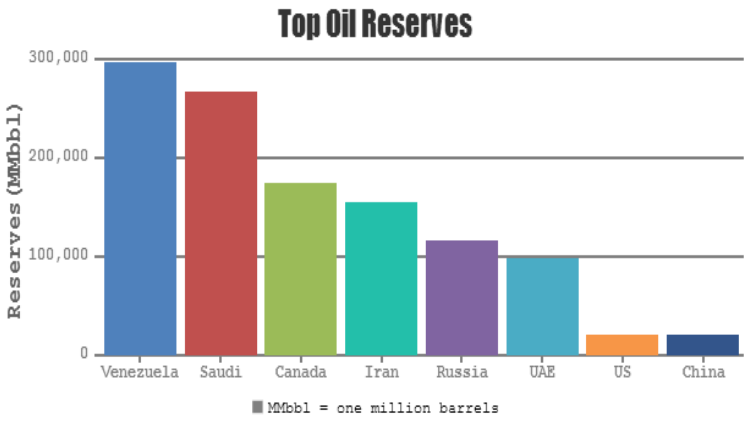


Fig 3: Bar chart

4. Line chart

A line chart or line graph is a type of chart which displays information as a series of data points called 'markers' connected by straight line segments. It is a basic type of chart common in many fields. A line chart is mostly used to represent a trend in data over intervals of time as shown in below figure.

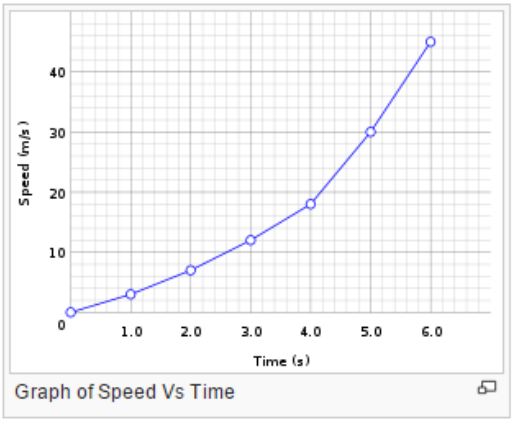


Fig 4: Line chart

5. Area chart

This is used to display quantitative data graphically. It is mainly based on the line chart. The area between axis and line are commonly emphasized with colors and textures. It uses the area chart for showing trends over time among related attributes as shown in below figure.

Fig 5: Area chart

6. Scatter plot

Scatter plot is graphical display of set of data in Cartesian coordinate, shows the relationship between two variables, one variable represent horizontal distance (independent variable) and second variable vertical distance (dependent variable) of data point from the coordinate axis [4]. Scatter plot shows the how strong the relationship are between the variables, and determines whether their exit any outlier in the data or not. It is use to look how the data is dispersed. Example of plot is shown in the following figure [5].

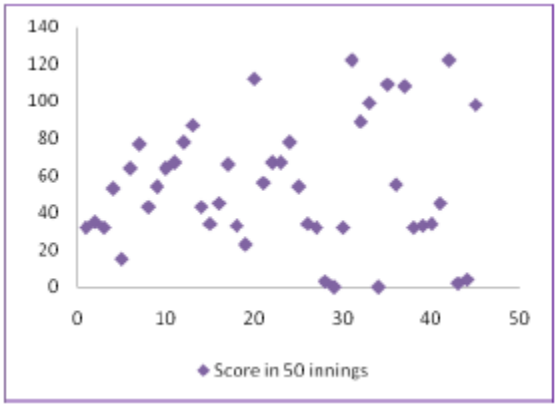


Fig 6: Scatter plot

III. POWER FEATURES IN EXCEL

There are many more basic techniques to represent data in excel like tree map, Entity relationship diagram, flow charts, venn diagrams, etc. These kind of techniques help user understand the data in quick way. Now, the focus is on massive data. Every field is now coming with big data. Many tools and techniques have been developed to process and analyze this big data. Excel was unable to give that much effective performance when analyzing and representing big data. It was getting slower to process the huge data to analyze and represent.

So, to overcome performance issues with massive data, Excel has been launched with 4 new features which are Power Pivot, Power query, Power view and power map. Power Pivot [6] is an Excel add-in that user can use to perform powerful data analysis and create sophisticated data models. With Power Pivot, user can mash up large volumes of data from various sources, perform information analysis rapidly, and share insights easily. Any data that is imported into Excel is available in Power Pivot, and vice versa. Power Query [7] for Excel can be found as an add-in in Excel 2013 which is used to simplify data discovery and allow easier access to different data sources. Power View [7] and Power Map are developed to improve the visualization of Microsoft Self-service BI. Power View can help users to create interactive reports and analytical views. Power Map enables the 3D map experience in Excel, with this feature, users can explore and navigate data in a particular geographical area.

1. Power Pivot

By using Power Pivot for Excel, we can transform enormous quantities of data with better speed into meaningful information to get the answers of queries. With Power Pivot for Excel, we can import millions of rows of data from multiple data sources into a single Excel workbook, create relationships between heterogeneous data, create calculated columns and measures using formulas, build PivotTables and Pivot Charts, and then further analyze the data so that you can make timely business decisions all without requiring IT assistance [8]. It overcomes existing limitations of traditional techniques for massive data analysis and representation.

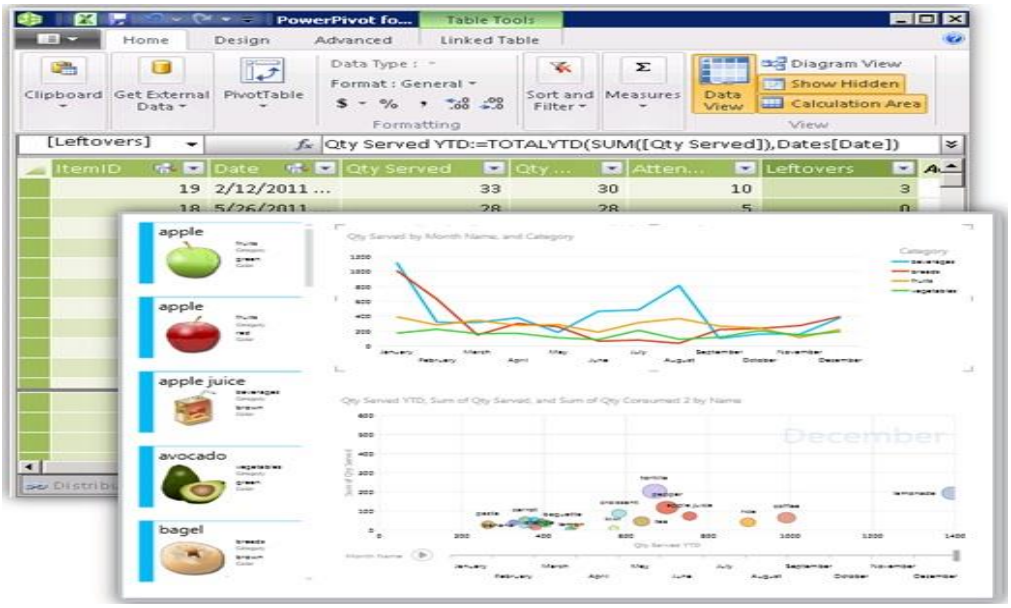


Fig 7: Reports with Power Pivot add-in [10]

As shown in above figure excel is with the add-in power pivot. It can be viewed in power pivot window by clicking on 'Manage'. By this, a separate window of power pivot will be opened that provide all the features i.e. formatting, sorting, filtering, creating relationships between tables, etc. We can add data from multiple sources by its button Get External Data. Then data can be shaped, analyzed and visualized.

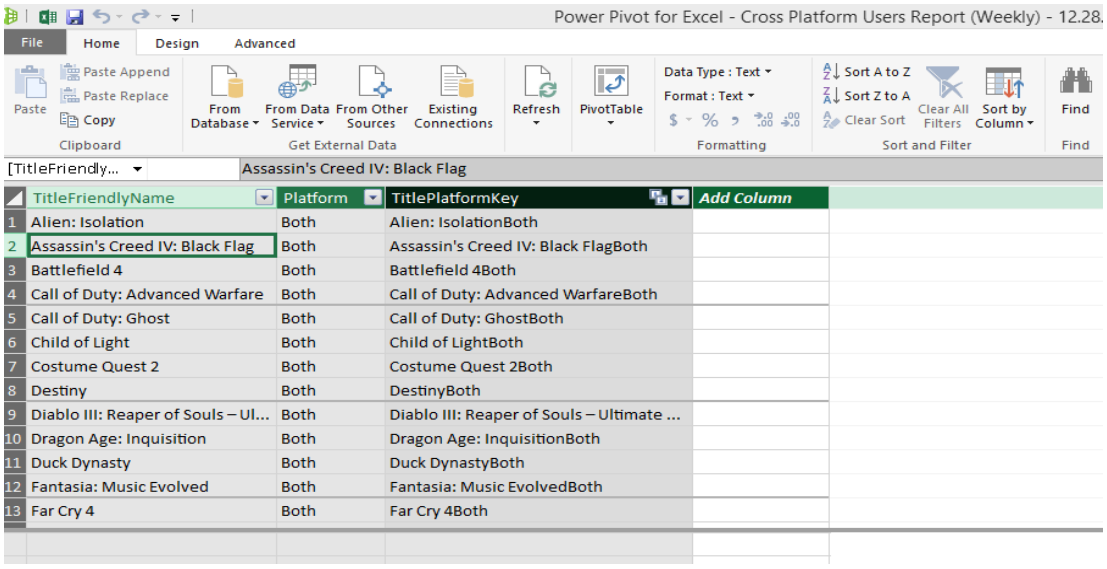


Fig 8: Power Pivot window

2. Limitations

Although Power Pivot feature provide far better performances than traditional techniques with massive amount of data, it has some limitations that need to be overcome.

1. File size of excel worksheet increase: When huge amount of data is loaded in power pivot, size of the excel worksheet increases considerably. As shown in below figure, file size is 194MB for 2,576,500 records of data. So, it is not much suitable to have this much size for bigger data.

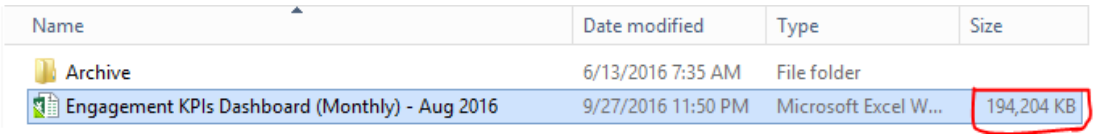


Fig 9: File size increase

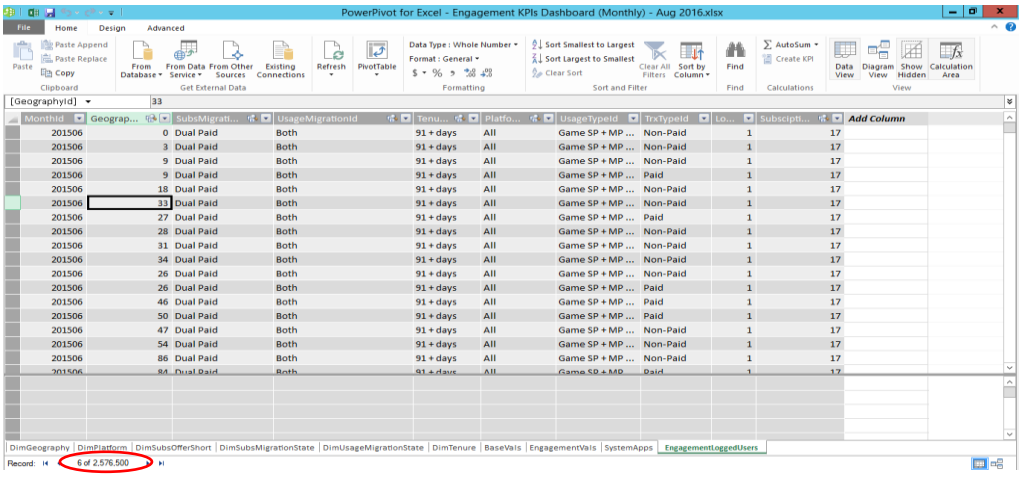


Fig 10: Bigger Data

2. Data Refreshes stuck at some places: Due to heavy data, power pivot cannot refresh all tabs in one shot because it stuck. And sometimes excel also crashes. So, we have to refresh all tabs sequentially which is time consuming.

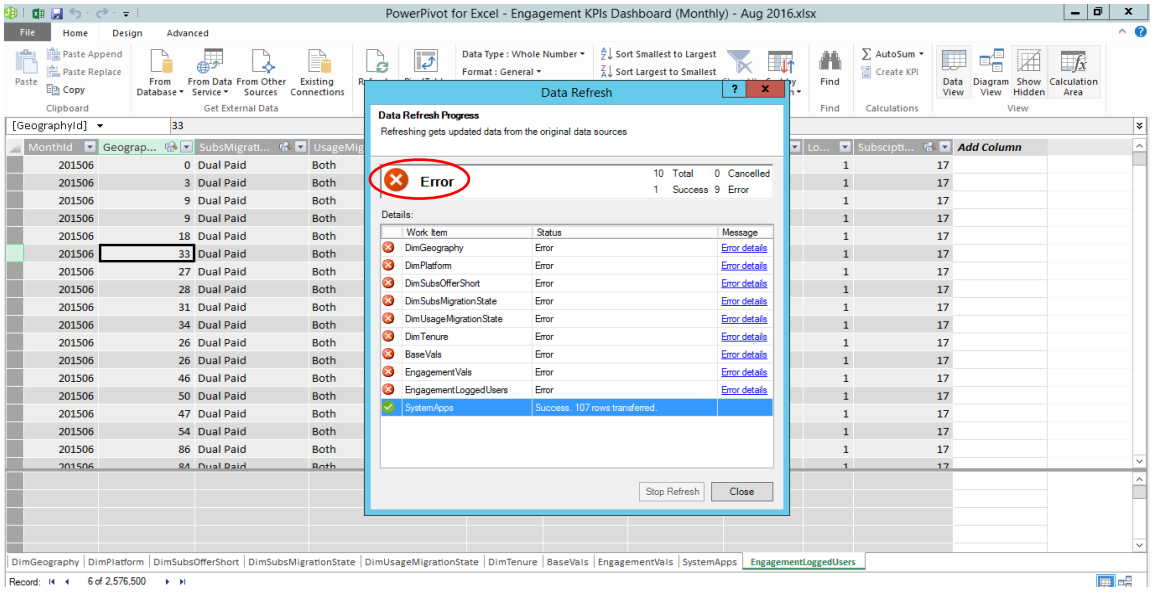


Fig 11: Data Refresh stuck

3. Sometimes, it doesn't respond quickly when too heavy data is loaded i.e. gives slow performance.

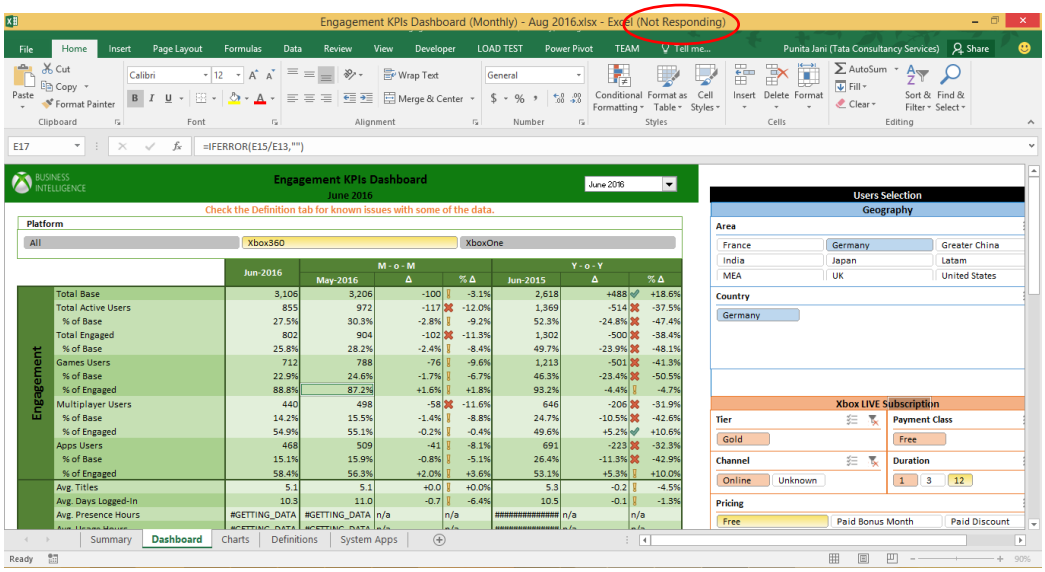


Fig 12: Not Responding

4. Basically, entire massive data is loaded to power pivot permanently so that it acquires memory

IV. CONCLUSION AND FUTURE SCOPE

There are number of tools for massive data visualizations in market. All these uses basic traditional approaches to import or dump all of the data to respective tool. In this paper, Excel has been described with its new features which are very effective but having some limitations. Although all these conventional visualization techniques are very useful for understanding the data, there is a need for new methods and approaches to improve the big data visualization scenario.

REFERENCES

[1] M. Khan, S.S. Khan, "Data and Information Visualization Methods and Interactive Mechanisms: A Survey", International Journal of Computer Applications, vol. 34, no. 1, 2011, pp. 1-14.

[2] C.L. P. Chen, C.-Y. Zhang, "Data-intensive applications, challenges, techniques and technologies: A survey on Big Data", Information Sciences, 275 (10), August 2014, pp. 314-347.

[3] JavaScript Column Charts | canvasjs. CanvasJs. N.p.,2016. http://canvasjs.com/html5-javascript-column-chart/. Accessed: 2016-10-01.

[4] Utts, Jessica M. (2005). "Seeing Through Statistics", 3rd Edition, Thomson Brooks/Cole, 2005, pp 166-167

[5] What is Scatter plot? : 2009. www.psychwiki.com/wiki/what_is_a_scatterplot%3F. Accessed: 2016-10-02.

[6] Power Pivot: Powerful data analysis and data modeling in Excel: 2016. https://support.office.com/en-us/article/Power-Pivot-Powerful-data-analysis-and-data-modeling-in-Excel-D7B119ED-1B3B-4F23-B634-445AB141B59B. Accessed: 2016-10-02.

[7] Mengwei Lu (2014), "Discovering Microsoft Self-service BI solution: Power BI", Haaga-Helia University of Applied Sciences, 2014, pp 1-94

[8] Power Pivot Add-in: 2016. https://support.office.com/en-us/article/Power-Pivot-Add-in-a9c2c6e2-cc49-4976-a7d7-40896795d045. Accessed: 2016-10-02

[9] Power Pivot: Powerful data analysis and data modeling in Excel: 2016. https://support.office.com/en-us/article/Power-Pivot-Powerful-data-analysis-and-data-modeling-in-Excel-D7B119ED-1B3B-4F23-B634-445AB141B59B. Accessed: 2016-10-02.

[10] Optimize a Sample PowerPivot Model for Power View Reports: 2016. https://technet.microsoft.com/en-us/library/hh759324(v=sql.110).aspx. Accessed: 2016-10-02.