

# The datapine Blog

News, Insights and Advice for Getting your Data in Shape

## Misleading Statistics Examples – Discover The Potential For Misuse of Statistics & Data In The Digital Age

By Mona Lebiad in Data Analysis, Aug 8th 2018



*“There are three types of lies – lies, damn lies, and statistics.” – Benjamin Disraeli*

Statistical analyses have historically been a stalwart of the high tech and advanced business industries, and today they are more important than ever. With the rise of advanced technology and globalized operations, statistical analyses grant businesses an insight into solving the extreme uncertainties of the market. Studies foster informed decision-making, sound judgments and actions carried out on the weight of evidence, not assumptions.

As businesses are often forced to follow a difficult-to-interpret market road map, statistical methods can help with the planning that is necessary to navigate a landscape filled with potholes, pitfalls and hostile competition. Statistical studies can also assist in the marketing of goods or services, and in understanding each target markets unique value drivers. In the digital age, these capabilities are only further enhanced and harnessed through the implementation of advanced technology and [business intelligence software](#). If all this true, what is the problem with statistics?

Actually, there is no problem per se – but there can be. Statistics are infamous for their ability and potential to exist as misleading and [bad data](#).

**Exclusive Bonus Content:** [Download Our Free Data Integrity Checklist](#)

Get our free checklist on ensuring data collection and analysis integrity!

### What Is A Misleading Statistic?

#### Search Blog

#### FOLLOW DATAPINE

[f](#) [t](#) [in](#)

#### Recent Posts

##### Top 10 Analytics And Business Intelligence Trends For 2022

Dec 2nd 2021

Check out what BI trends will be on everyone's lips and keyboards in 2022.

[read more](#)

##### Why Data Driven Decision Making is Your Path To Business Success

Nov 24th 2021

Data means intelligence, and by using it to your advantage, you will accelerate your business success. Here, we explain everything you need to know about data driven decisions.

[read more](#)

##### Utilize The Effectiveness Of Professional Executive Dashboards & Reports

Nov 18th 2021

The unrivaled power and potential of executive dashboards, metrics and reporting explained.

[read more](#)

#### Tags / Categories

- [Business Intelligence](#) (46)

something wrong if he or she does not notice the error or the does not have the full data picture.

Given the importance of data in today’s rapidly evolving digital world, it is important to be familiar with the basics of misleading statistics and oversight. As an exercise in due diligence, we will review some of the most common forms of misuse of statistics, and various alarming (and sadly, common) misleading statistics examples from public life.

Are Statistics Reliable?

73.6% of statistics are false. Really? No, of course it’s a made-up number (even though such a study would be interesting to know – but again, could have all the flaws it tries at the same time to point out). Statistical reliability is crucial in order to ensure the precision and validity of the analysis. To make sure the reliability is high, there are various techniques to perform – first of them being the control tests, that should have similar results when reproducing an experiment in similar conditions. These controlling measures are essential and should be part of any experiment or survey – unfortunately, that isn’t always the case.

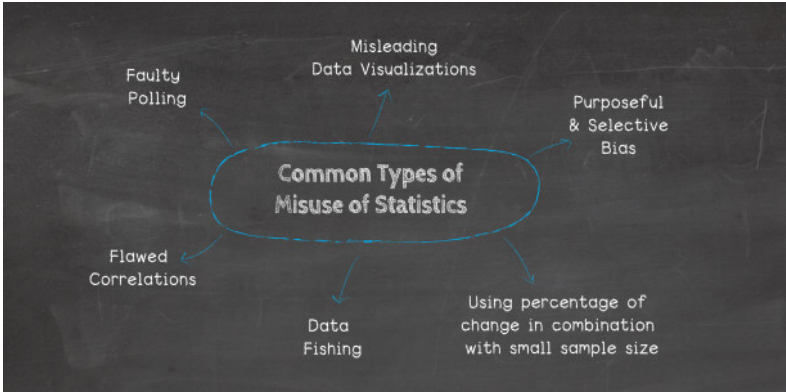
While numbers don’t lie, they can in-fact be used to mislead with half-truths. This is known as the “misuse of statistics.” It is often assumed that the misuse of statistics is limited to those individuals or companies seeking to gain profit from distorting the truth, be it economics, education or mass media.

However, the telling of half-truths through study is not only limited to mathematical amateurs. A 2009 investigative survey by Dr. Daniele Fanelli from The University of Edinburgh found that 33.7% of scientists surveyed admitted to questionable research practices, including modifying results to improve outcomes, subjective data interpretation, withholding analytical details and dropping observations because of gut feelings.... Scientists!

While numbers don’t always have to be fabricated or misleading, it is clear that even societies most trusted numerical gatekeepers are not immune to the carelessness and bias that can arise with statistical interpretation processes. There are different ways how statistics can be misleading that we will detail later. The most common one is of course correlation versus causation, that always leaves out another (or two or three) factor that are the actual causation of the problem. Drinking tea increases diabetes by 50%, and baldness raises the cardiovascular disease risk up to 70%! Did we forget to mention the amount of sugar put in the tea, or the fact that baldness and old age are related – just like cardiovascular disease risks and old age?

So, can statistics be manipulated? They sure can. Do numbers lie? You can be the judge.

How Statistics Can Be Misleading



- [Data Analysis](#) (51)
- [Data Visualization](#) (32)
- [Guest Articles](#) (18)
- [KPIs](#) (27)
- [Market Intelligence](#) (1)
- [News](#) (3)
- [Real World Stats](#) (10)
- [Reporting](#) (46)
- [Technology](#) (6)

PRODUCT

SOLUTIONS

SERVICES

LEARN

FREE

necessary to create misunderstandings. The misuse of statistics is a much broader problem that now permeates through multiple industries and fields of study. Here are a few potential mishaps that commonly lead to misuse:

- **Faulty polling**

The manner in which questions are phrased can have a huge impact on the way an audience answers them. Specific wording patterns have a persuasive effect and induce respondents to answer in a predictable manner. For example, on a poll seeking tax opinions, let's look at the two potential questions:

- Do you believe that you should be taxed so other citizens don't have to work? - Do you think that the government should help those people who cannot find work?

These two questions are likely to provoke far different responses, even though they deal with the same topic of government assistance. These are examples of "loaded questions."

A more accurate way of wording the question would be, "Do you support government's assistance programs for unemployment?" or, (even more neutrally) "What is your point of view regarding unemployment assistance?"

The latter two examples of the original questions eliminate any inference or suggestion from the poller, and thus, are significantly more impartial. Another unfair method of polling is to ask a question, but precede it with a conditional statement or a statement of fact. Staying with our example, that would look like this: "Given the rising costs to the middle class, do you support government assistance programs?"

A good rule of thumb is to always take polling with a grain of salt, and to try to review the questions that were actually presented. They provide great insight, often more so than the answers.

- **Flawed correlations**

The problem with correlations is this: if you measure enough variables, eventually it will appear that some of them correlate. As **one out of twenty** will inevitably be deemed significant without any direct correlation, studies can be manipulated (with enough data) to prove a correlation that does not exist or that is not significant enough to prove causation.

To illustrate this point further, let's assume that a study has found a correlation between an increase in car accidents in the state of New York in the month of June (A), and an increase in bear attacks in the state of New York in the month of June (B).

That means there will likely be six possible explanations:

- Car accidents (A) cause bear attacks (B) - Bear attacks (B) cause car accidents (A) - Car accidents (A) and bear attacks (B) partly cause each other - Car accidents (A) and bear attacks (B) are caused by a third factor (C) - Bear attacks (B) are caused by a third factor (C) which correlates to car accidents (A) - The correlation is only chance

Any sensible person would easily identify the fact that car accidents do not cause bear attacks. Each is likely a result of a third factor, that being: an increased population, due to high tourism season in the month of June. It would be preposterous to say that they cause each other... and that is exactly why it is our example. It is easy to see a correlation.

But, what about causation? What if the measured variables were different? What if it was something more believable, like Alzheimer's and old age? Clearly there is a correlation between the two, but is there causation? Many would falsely assume, yes,

PRODUCT

SOLUTIONS

SERVICES

LEARN

FREE

COMPANY

- **Data fishing**

This misleading data example is also referred to as “data dredging” (and related to flawed correlations). It is a data mining technique where extremely large volumes of data are analyzed for the purposes of discovering relationships between data points. Seeking a relationship between data isn't a data misuse per se, however, doing so without a hypothesis is.

Data dredging is a self-serving technique often employed for the unethical purpose of circumventing traditional data mining techniques, in order to seek additional data conclusions that do not exist. This is not to say that there is no proper use of data mining, as it can in-fact lead to surprise outliers and interesting analyses. However, more often than not, data dredging is used to assume the existence of data relationships without further study.

Often times, data fishing results in studies that are highly publicized due to their important or outlandish findings. These studies are [very soon contradicted](#) by other important or outlandish findings. These false correlations often leave the general public very confused, and searching for answers regarding the significance of causation and correlation.

Likewise, another common practice with data is the omission, meaning that after looking at a large data set of answers, you only pick the ones that are supporting your views and findings and leave out those that contradict it. As mentioned in the beginning of this article, it has been shown that a third of the scientists admitted that they had questionable research practices, including withholding analytical details and modifying results...! But then again, we are facing a study that could itself fall into these 33% of questionable practices, faulty polling, selective bias... It becomes hard to believe any analysis!

**Exclusive Bonus Content:** [Download Our Free Data Integrity Checklist](#)

Get our free checklist on ensuring data collection and analysis integrity!

- **Misleading data visualization**

Insightful graphs and charts include very basic, but essential, grouping of elements. Whatever the [types of data visualization](#) you choose to use, it must convey:

- The scales used - The starting value (zero or otherwise) - The method of calculation (e.g., dataset and time period)

Absent these elements, visual data representations should be viewed with a grain of salt, taking into account the common [data visualization mistakes](#) one can make. Intermediate data points should also be identified and context given if it would add value to the information presented. With the increasing reliance on intelligent solution automation for variable data point comparisons, best practices (i.e., design and scaling) should be implemented prior to comparing data from different sources, datasets, times and locations.

- **Purposeful and selective bias**

The last of our most common examples for misuse of statistics and misleading data is, perhaps, the most serious. Purposeful bias is the deliberate attempt to influence data findings without even feigning professional accountability. Bias is most likely to take the form of data omissions or adjustments.

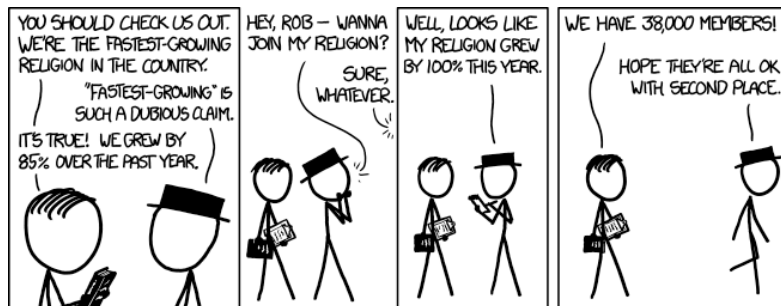
The selective bias is slightly more discreet for whom does not read the small lines. It usually falls down on the sample of people surveyed. For instance, the nature of the

statistical error called “selective bias”. COMPANY

- **Using percentage change in combination with a small sample size**

Another way of creating misleading statistics, also linked with the choice of sample discussed above, is the size of said sample. When an experiment or a survey is led on a totally not significant sample size, not only will the results be unusable, but the way of presenting them - namely as percentages - will be totally misleading.

Asking a question to a sample size of 20 people, where 19 answer "yes" (=95% say for yes) versus asking the same question to 1,000 people and 950 answer "yes" (=95% as well): the validity of the percentage is clearly not the same. Providing solely the percentage of change without the total numbers or sample size will be totally misleading. [xkcd's comic](#) illustrate this very well, to show how the "fastest-growing" claim is a totally relative marketing speech:

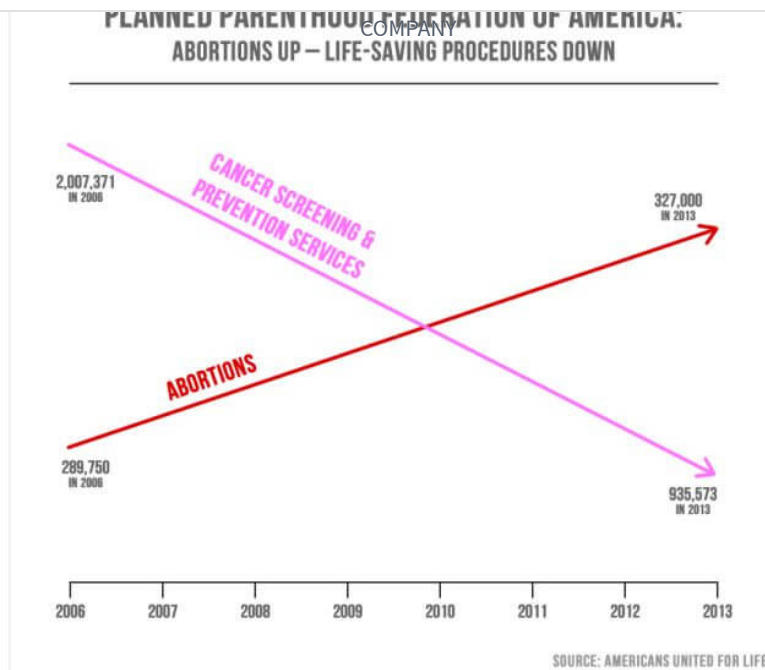


Likewise, the needed sample size is influenced by the kind of question you ask, the statistical significance you need (clinical study vs business study), and the statistical technique. If you perform a quantitative analysis, sample sizes under 200 people are usually invalid.

## Misleading Statistics Examples In Real Life

Now that we have reviewed several of the most commons methods of data misuse, let's look at various digital age examples of misleading statistics across three distinct, but related, spectrums: media and politics, advertising and science. While certain topics listed here are likely to stir emotion depending on one's point of view, their inclusion is for data demonstration purposes only.

- **Examples of misleading statistics in the media and politics**

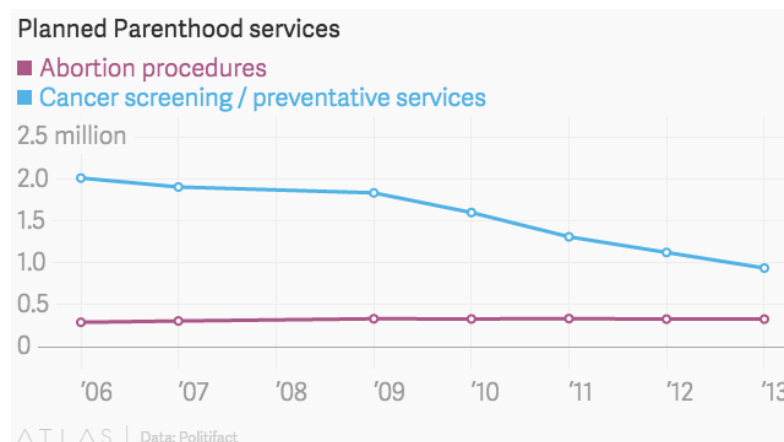


Misleading statistics in the media are quite common. On Sept. 29, 2015, Republicans from the U.S. Congress questioned Cecile Richards, the president of Planned Parenthood, regarding the misappropriation of \$500 million in annual federal funding. The above graph/chart was presented as a point of emphasis.

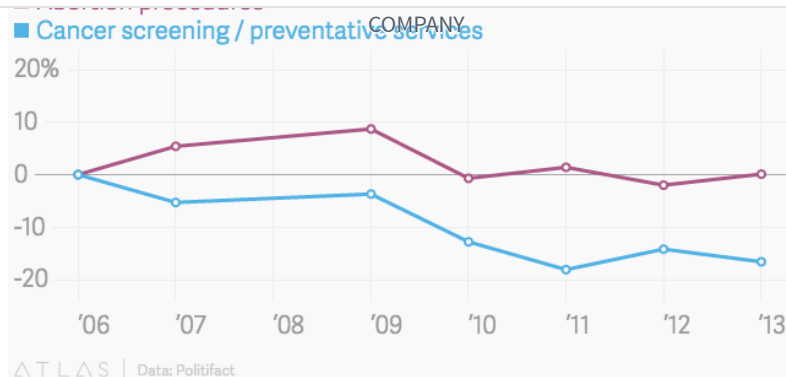
Representative Jason Chaffetz of Utah explained: “In pink, that’s the reduction in the breast exams, and the red is the increase in the abortions. That’s what’s going on in your organization.”

Based on the structure of the chart, it does in-fact appear to show that the number of abortions since 2006 experienced substantial growth, while the number of cancer screenings substantially decreased. The intent is to convey a shift in focus from cancer screenings to abortion. The chart points appear to indicate that 327,000 abortions are greater in inherent value than 935,573 cancer screenings. Yet, closer examination will reveal that the chart has no defined y-axis. This means that there is no definable justification for the placement of the visible measurement lines.

PolitiFact, a fact checking advocacy website, [reviewed](#) Rep. Chaffetz’s numbers via a comparison with Planned Parenthood’s own annual reports. Using a clearly defined scale, here is what the information looks like:



And like this with another valid scale:



Once placed within a clearly defined scale, it becomes evident that while the number of cancer screenings has in fact decreased, it still far outnumbers the quantity of abortion procedures performed yearly. As such, this is a great misleading statistics example, and some could argue bias considering that the chart originated not from the Congressman, but from Americans United for Life, an anti-abortion group. This is just one of many examples of misleading statistics in the media and politics.

- Misleading statistics in advertising



In 2007, Colgate was ordered by the Advertising Standards Authority (ASA) of the U.K. to abandon their claim: “More than 80% of Dentists recommend Colgate.” The slogan in question was positioned on an advertising billboard in the U.K., and was deemed to be in breach of U.K. advertising rules.

The claim, which was based on surveys of dentists and hygienists carried out by the manufacturer, was found to be misrepresentative as it allowed the participants to select one or more toothpaste brands. The ASA stated that the claim “... would be understood by readers to mean that 80 percent of dentists recommend Colgate over and above other brands, and the remaining 20 percent would recommend different brands.”

The ASA continued, “Because we understood that another competitor’s brand was recommended almost as much as the Colgate brand by the dentists surveyed, we concluded that the claim misleadingly implied 80 percent of dentists recommend Colgate toothpaste in preference to all other brands.” The ASA also claimed that the scripts used for the survey informed the participants that the research was being performed by an independent research company, which was inherently false.



would fall under faulty polling and outright lies.

**Exclusive Bonus Content:** [Download Our Free Data Integrity Checklist](#)

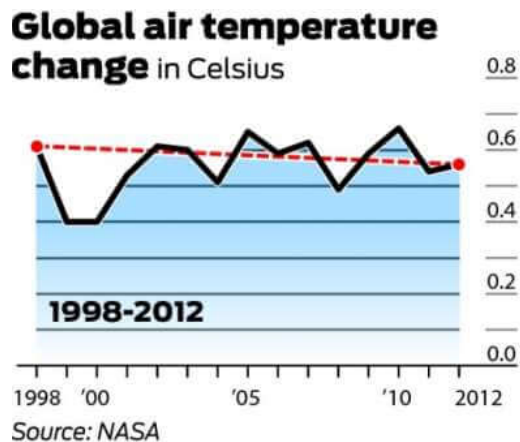
Get our free checklist on ensuring data collection and analysis integrity!

- **Misleading statistics in science**

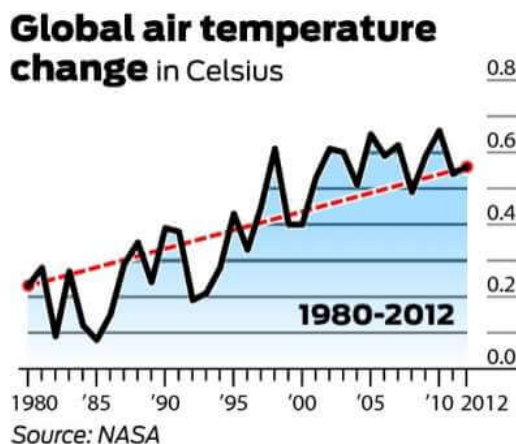
Much like abortion, global warming is another politically charged topic that is likely to arouse emotions. It also happens to be a topic that is vigorously endorsed by both opponents and proponents via studies. Let's take a look at some of the evidence for and against.

It is generally agreed upon that the global mean temperature in 1998 was 58.3 degrees Fahrenheit. This is according to [NASA's Goddard Institute for Space Studies](#). In 2012, the global mean temperature was measured at 58.2 degrees. It is, therefore, argued by global warming opponents that, as there was a 0.1 degree decrease in the global mean temperature over a 14-year period, global warming is disproved.

The below graph is the one most often referenced to disprove the global warming. It demonstrates the change in air temperature (Celsius) from 1998 to 2012.

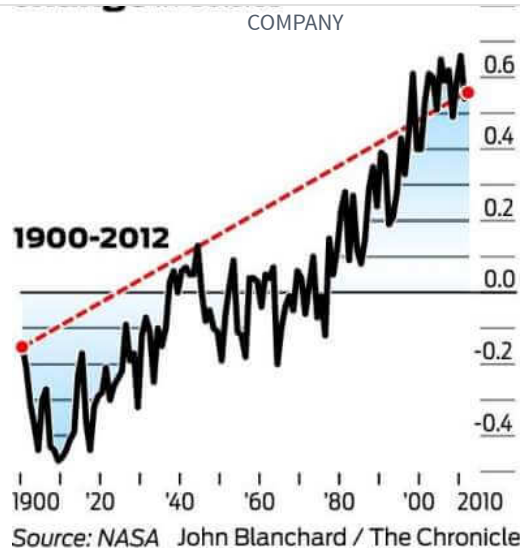


It is worth mentioning that 1998 was one of the hottest years on record due to an abnormally strong El Niño wind current. It is also worth noting that, as there is a large degree of variability within the climate system, temperatures are typically measured with at least a 30-year cycle. The below chart expresses the 30-year change in global mean temperatures.



And now have a look at the trend from 1900 to 2012:





While the long-term data may appear to reflect a plateau, it clearly paints a picture of gradual warming. Therefore, using the first graph, and only the first graph, to disprove global warming is a perfect misleading statistics example.

## How To Read Statistics With Distance

A first good thing would be of course to stand in front an honest survey/experiment/research – pick the one you have beneath your eyes –, that has applied the correct techniques of collection and [interpretation of data](#). But you cannot know until you ask yourself a couple of questions and analyze the results you have in between your hands.

As entrepreneur and former consultant Mark Suster [advises in an article](#), you should wonder who did the primary research of said analysis. Independent university study group, lab-affiliated research team, consulting company? From there naturally stems out the question: who paid them? As no one works for free, it is always interesting to know who sponsors the research. Likewise, what are the motives behind the research? What did the scientist or statisticians tried to figure out? Finally, how big was the sample set and who was part of it? How inclusive was it?

These are important questions to ponder and answer before spreading everywhere skewed or biased results – even though it happens all the time, because of amplification. A typical example of amplification happens often with newspapers and journalists, who take one piece of data and need to turn it into headlines – thus often out of its original context. No one buys a magazine where it states that next year, the same thing is going to happen in XYZ market as this year – even though it is true. Editors, clients, and people want something new, not something they know; that's why we often end up with an amplification phenomenon that gets echoed and more than it should.

## Misuse of Statistics - A Summary

To the question "can statistics be manipulated?", we can address 6 methods often used - on purpose or not - that skew the analysis and the results. Here are common types of misuse of statistics:

- Faulty polling
- Flawed correlations
- Data fishing
- Misleading data visualization
- Purposeful and selective bias

PRODUCT

SOLUTIONS

SERVICES

LEARN

FREE

Now that you know them, it will be easier to spot them out and question all the stats that are given to you every day. Likewise, in order to ensure you keep a certain distance to the studies and surveys you read, remember the questions to ask yourself - who researched and why, who paid for it, what was the sample.

**Exclusive Bonus Content:** [Download Our Free Data Integrity Checklist](#)

Get our free checklist on ensuring data collection and analysis integrity!

## Transparency and Data-Driven Business Solutions

While it is quite clear that statistical data has the potential to be misused, it can also ethically drive market value in the digital world. Big data has the ability to provide digital age businesses with a roadmap for efficiency and transparency, and eventually, profitability. Advanced technology solutions like [online reporting software](#) can enhance statistical data models, and provide digital age businesses with a step-up on their competition.

Whether for market intelligence, customer experience or business reporting, the future of data is now. Take care to apply data responsibly, ethically and visually, and watch your transparent corporate identity grow.

**FOLLOW DATAPINE**



Managing Partners: Martin Blumenau, Jakob Reherrmann | Trade Register:  
Berlin-Charlottenburg HRB 144962 B | Tax Identification Number: DE 28 552 2148

FRIEDELSTRASSE 27  
12047 BERLIN, GERMANY  
+49 30 6920 9427  
[CONTACT@DATAPINE.COM](mailto:CONTACT@DATAPINE.COM)