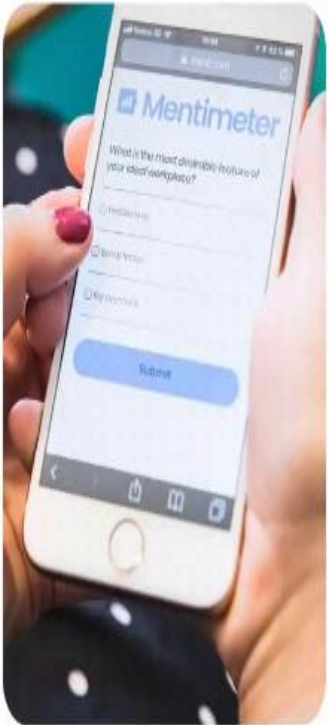


# Data Mining and Data Analytics

- ❖ Data Mining
- ❖ Data Analytics in Balance Sheet

**CA Gaurav Mishra**  
**+91 - 8108763725**

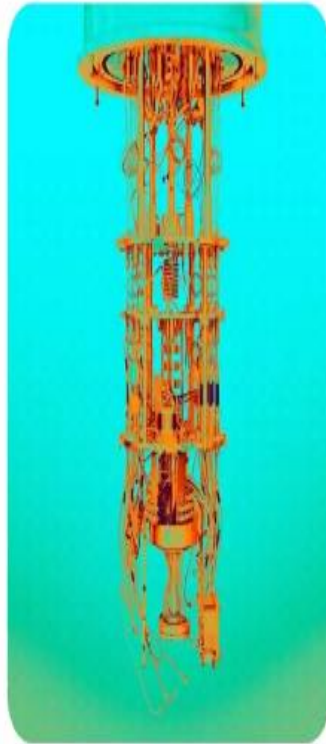
# Objective of the Session



MENTIMETER  
QUESTIONS



INTRODUCTION TO  
DATA SCIENCE



PROBLEMS YOU  
CAN SOLVE



USE CASES FOR  
THE SUPER 7





Stakeholders  
Perspective



Benefits of Data  
Analysis

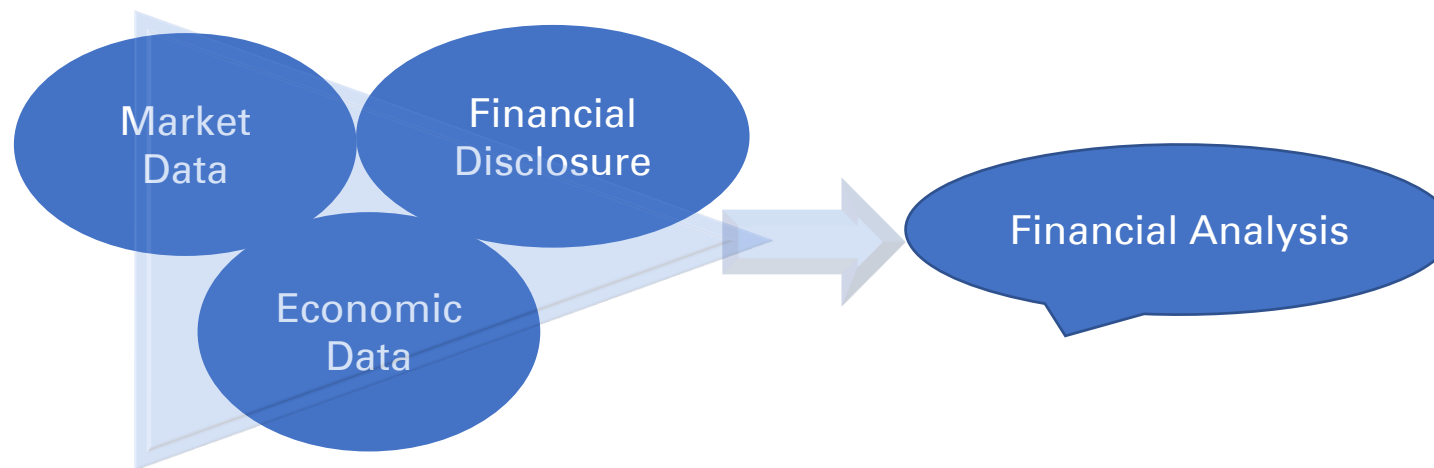
# Agenda for Discussion

## Agenda

Topic	Description
<b>Data Mining and Data Analytics</b>	
Financial Statement	<ul style="list-style-type: none"><li>▪ About Financial Statement and its Key Objective</li><li>▪ Stakeholders of Financial Statement</li><li>▪ Key KPI's monitored by various stakeholders of Financial Statements</li></ul>
 Data Mining and Data Analysis	<ul style="list-style-type: none"><li>▪ About Data Mining and Data analytics</li><li>▪ Benefits / Impact / Limitation of Data Analysis</li><li>▪ Data Analysis Tools available in market</li><li>▪ Way forward of Data Analysis</li></ul>
 Key Data Analysis to be performed	<ul style="list-style-type: none"><li>▪ What Data Analysis can be performed in Financial Statement</li><li>▪ Key Ratios and perspective of stakeholders of financial statement</li><li>▪ Recent trend of Presentation of financial information to various stakeholders</li></ul>

# Financial Statement

- ❖ **Financial Statement means,** Written records reports prepared by a company's management to present the financial performance and position at a point in time.
- ❖ **Financial Statement Analysis,**
  - ❑ Process of selecting, evaluating, and interpreting financial data,
  - ❑ Along with other pertinent information, in order to formulate an assessment of company's present and future financial condition and performance
- ❑ 'Analysis' means,
  - ❑ simplification of financial data by methodical classification of the data given in the financial statements,
  - ❑ 'interpretation' means, 'explaining the meaning and significance of the data so simplified.'
- ❑ However, both 'analysis and interpretation' are interlinked and complementary to each other.



# Financial Statement - Component

Statement of  
Equity and Liability

Statement of Profit  
and Loss

Cash Flow  
Statement

Statement of  
Changes in Equity

Notes to Financial  
Statement

# Stakeholders of Financial Statement



# KPI's of Financial Statement for various stakeholder

## Key Performance Indicator

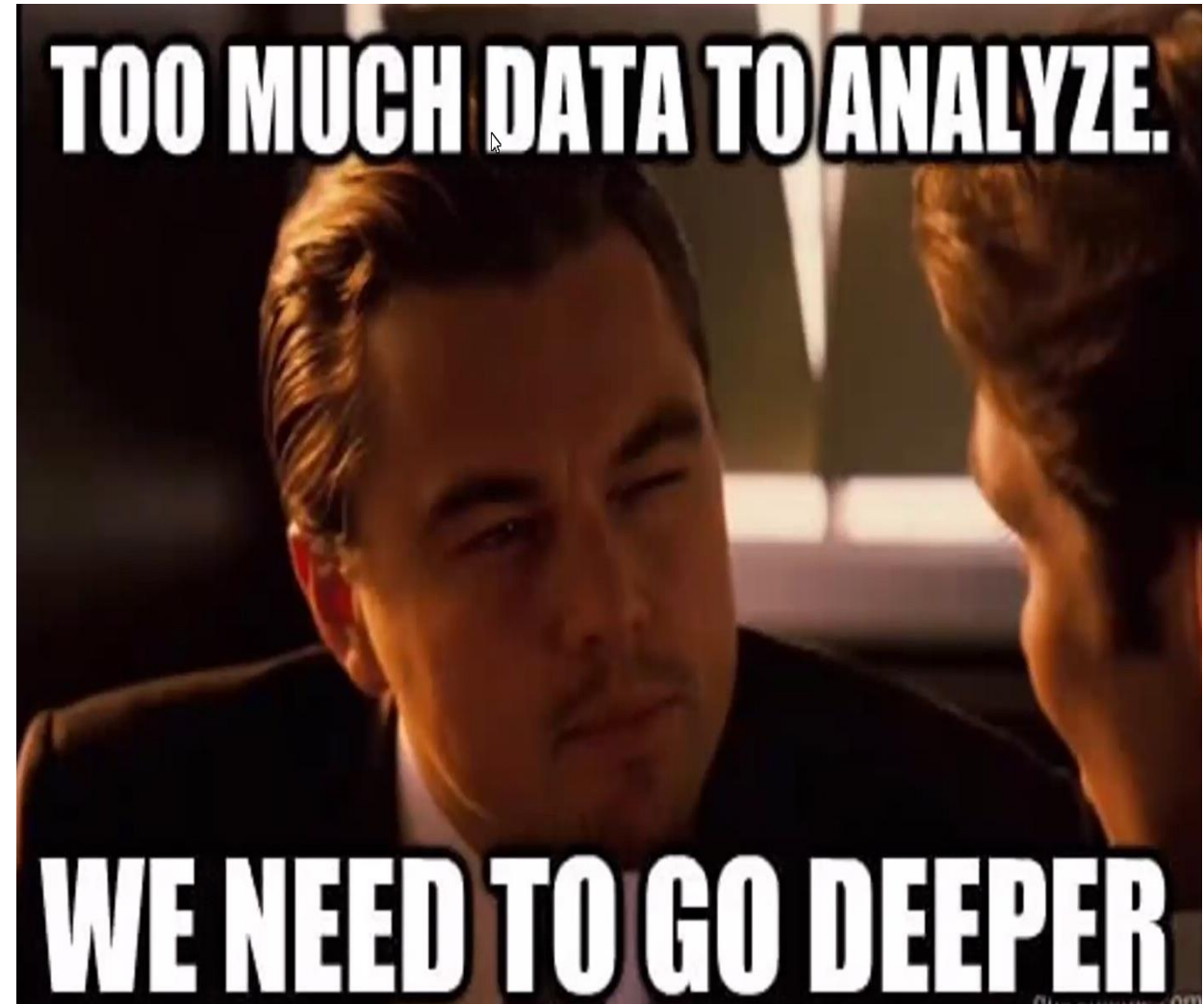
1. **Management:** Management of the company looks at the financial statement from the perspective of liquidity, profitability, cash flows, internal control and future prospect to take financial decision.
2. **Investors:** Like to make the decision based on the financial statement whether they need to keep invested or move out of the company based on its performance.
3. **Customers:** Big clients would like to have a long-term partnership or contract with the company thus they would like to work with a company that is financially stable.
4. **Suppliers:** Judging the probability of firm's continued ability to meet all its financial obligations in the future.
5. **Competitors:** Decide to change their strategy looking at the statements.
6. **Government and Government Agencies:** would like to do future tax predictions based on the performance of the company and industry practices.
7. **Employees:** they would look to have a deep understanding of the business and the current industry situation which will be available in the financial statements.
8. **Lenders:** For determining a company's ability to generate cash, to pay interest and repay the principal amount



The background is a complex, abstract geometric pattern in various shades of blue and teal. It features a series of parallel, slightly curved lines that create a sense of depth and perspective, leading towards a bright, glowing light source in the center-right. The overall effect is futuristic and high-tech.

# Data Science

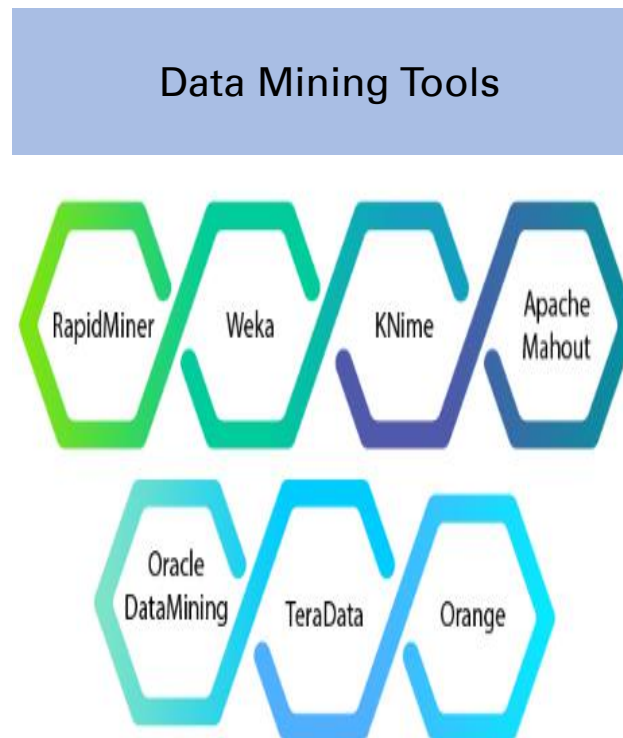
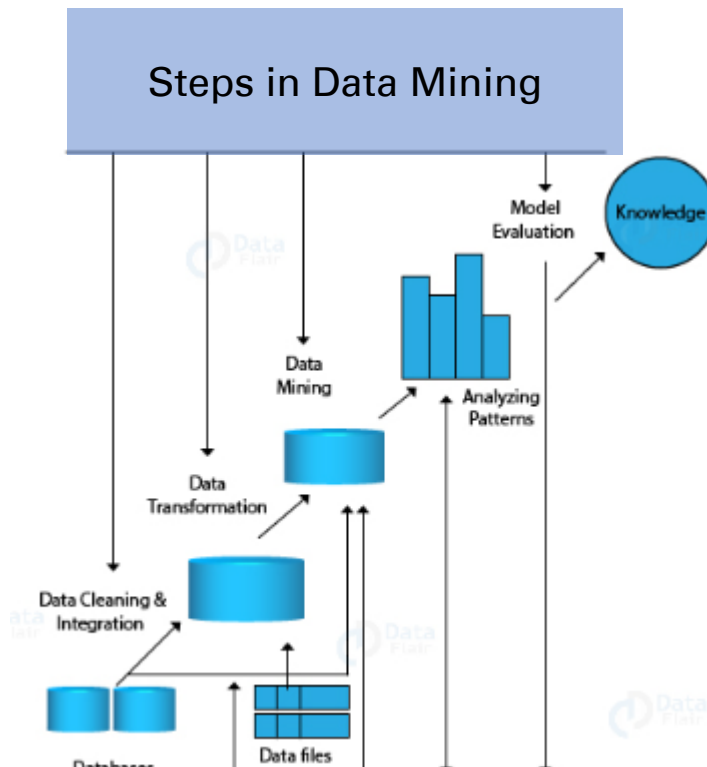




# Data Mining

## Data Mining

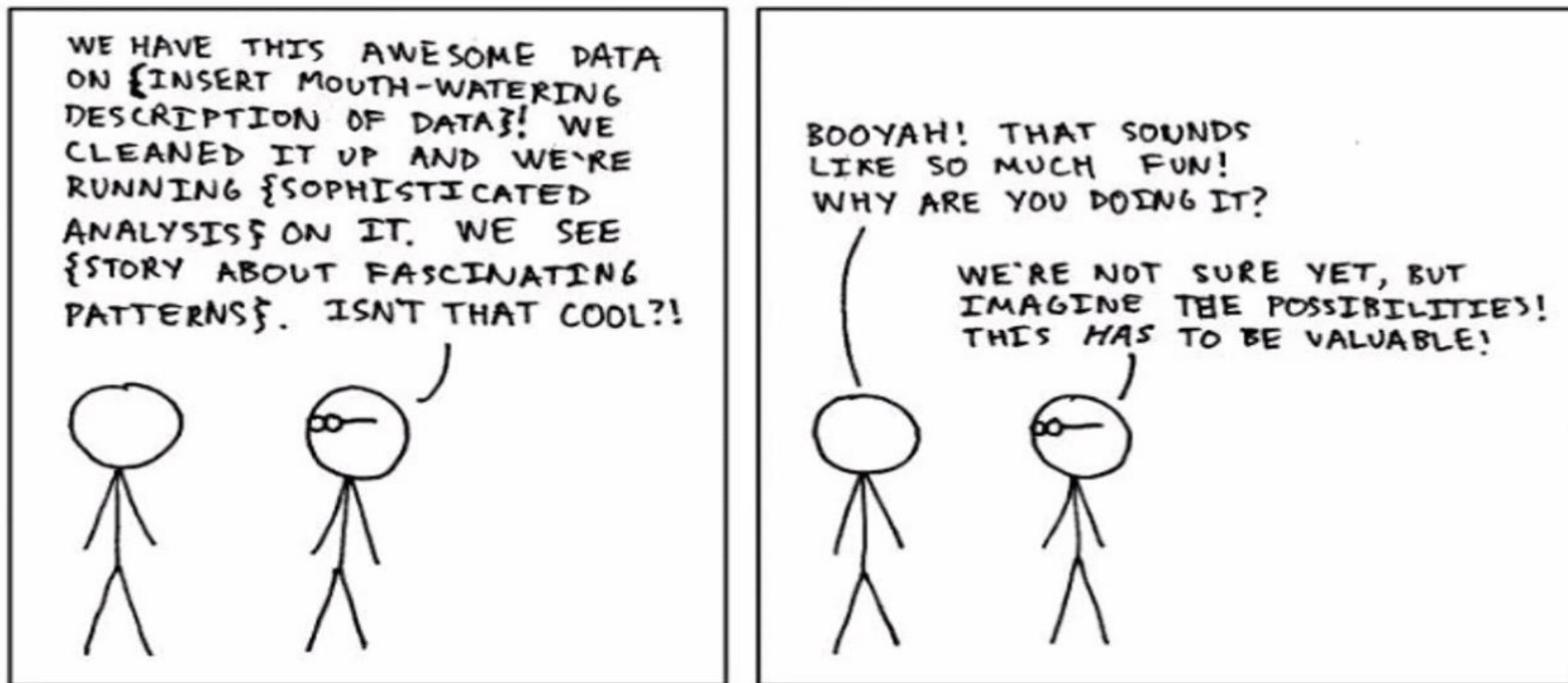
The process of data mining is a complex process that involves intensive data warehousing as well as powerful computational technologies. Furthermore, data mining is not only limited to the extraction of data but is also used for transformation, cleaning, data integration, and pattern analysis. Another terminology for **Data Mining is Knowledge Discovery**.



- Data Mining Applications**
- Market and Stock Analysis
  - Fraud Detection
  - Risk Management and Corporate Analysis
  - Analyzing the Customer Lifetime Value

## What is Data Science?

Data science is an inter-disciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from many structural and unstructured data.



## Importance of Data Science in Business



## Applications of Data Science in Finance





# Data Science

## Steps in Data Science



## Tools for Data Science

**Python**  
Programming language  
and software  
development

**R**  
visualizing and  
analyzing data

**SAS: Statistical  
Analysis System**

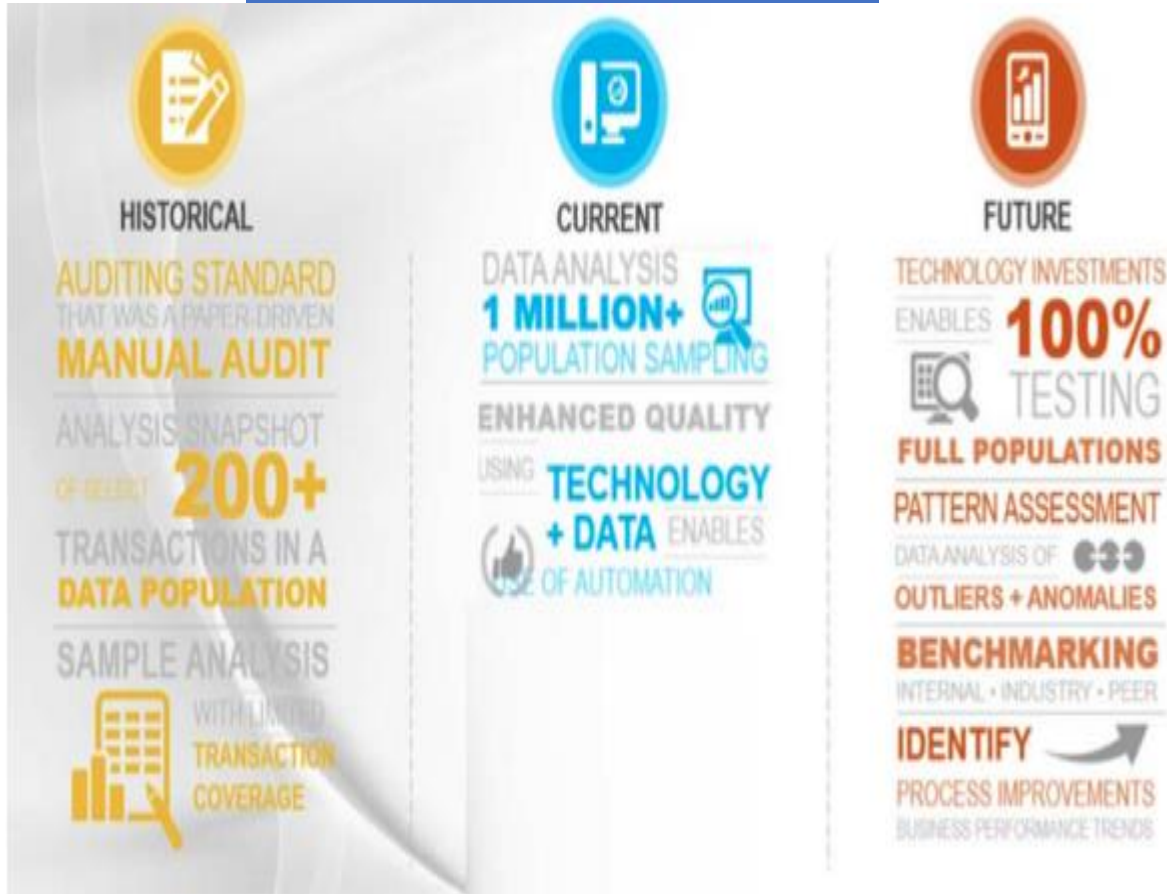
**Apache Spark:**  
Data processing and  
analysis capabilities

**D3.js:**  
javascript based  
library

**Tableau:**  
**Visualization  
Software**

# Data Analytics in Financial Statement

## Evolution of Audit



## Evolution of Big Data





# Data Analytics in Financial Statement

**“Data analytics** is the process of examining data sets in order to draw conclusions about the information they contain, increasingly with the aid of specialized systems and software.”

**“Big Data Analytics** is the process of inspecting, cleaning, transforming, and modeling Big Data to discover and communicate useful information and patterns, suggest conclusions, and support decision making.”

		Data Analytic Techniques	
		Traditional (Excel, ACL, Idea)	Extended (Visualization, Predictive analytics)
Data Sources	Traditional (Accounting & Financial)	A	B
	Extended (Non-Financial Data → Big Data)	C	D

# Data Analytics in Financial Statement

## **Audit with Big Data as opposed to audit of Big Data**

- Identifying and assessing the risks
- associated with accepting or continuing an audit engagement
- of material misstatement through understanding the entity and its environment
- of material misstatement of the financial statements due to fraud, and testing for fraud having regard to the assessed risks
- Performing analytical procedures
- In response to the auditor's assessment of the risks of material misstatement
- near the end of the audit to assist the auditor form an overall conclusion

## **Characteristics of Data Analytics**

- Population vs sample analysis
- Correlations as opposed to causation
- Integration with continuous auditing.

# Implementing Data Analytics in Audit

- Providers of Big Data analytics
- Exception prioritization (Issa and Kogan 2013)
- Messy data in audit
- Computational challenges
- Data subsetting
- Legal challenges and audit standards.

# Data Analytics Tools

Tool	Characteristics		
Excel	<ul style="list-style-type: none"> <li>- Ability to handle considerable volumes of data</li> <li>- Available on every desktop</li> <li>- Familiar user interface</li> <li>- Affordable and compatible</li> <li>- Powerful features (e.g., subtotals, lookups, pivots, etc.)</li> </ul> (Shah, 2004)	SAS	<ul style="list-style-type: none"> <li>- Analyzes any kind and size of data</li> <li>- Readily understands results with a wealth of graphs</li> <li>- Applies the latest statistical techniques</li> <li>- Uses proven and validated methods</li> </ul> (SAS, 2017)
IDEA	<ul style="list-style-type: none"> <li>- Analyzes large amounts of data</li> <li>- User-friendly interface (little technical skill required)</li> <li>- High speed of processing large amounts of transactions</li> <li>- Detects and handles fraud easily</li> </ul> (Bagga & Singh, 2001)		
ACL	<ul style="list-style-type: none"> <li>- Data is locked down as read-only</li> <li>- Commands are auditor-friendly</li> <li>- Automatically records all of the commands that are run and the results of the procedures in its log</li> <li>- Makes use of Writing Scripts (Batch feature)</li> </ul> (Dugas, s.d.)	Data Mining	<ul style="list-style-type: none"> <li>- Automatically discovers useful information even from complex data sets</li> <li>- Automatically discovers unknown patterns</li> <li>- Handles large amounts of data</li> <li>- Relatively high cost</li> <li>- Technical skills required</li> </ul> (Bagga & Singh, 2011)

# Changes within Audit due to Data Analysis

Major Changes	Characteristics
Collection of Evidence	<b>Challenges on Dealing with Information in a BD Environment</b> <ul style="list-style-type: none"> <li>- Relevance and reliability of external/internal data</li> <li>- Safeguarding data provenance</li> <li>- Acquisition of data</li> <li>- Information overload</li> <li>- Pattern recognition</li> <li>- Information relevance</li> <li>- Privacy and security issues</li> </ul> <b>Technological Changes</b> <ul style="list-style-type: none"> <li>- New innovative analytical tools: ACL, IDEA, SAS and Data Mining</li> <li>- Full population examination</li> <li>- BDA in all phases of the audit process</li> </ul>
Basic Concepts	<ul style="list-style-type: none"> <li>- Reduction of materiality</li> <li>- Impairment of the auditor's independence</li> <li>- Shift in method of judgment: from causation to correlation</li> </ul>

Timing of the Audit	<b>Continuous Auditing</b> <ul style="list-style-type: none"> <li>- Enhances audit quality and client service</li> <li>- Possibility to spread the workload throughout the year</li> <li>- Fewer chances for deviations to expand into larger levels of error</li> <li>- Not feasible for all entities due to limitations in data quality, quantity and accessibility</li> <li>- Gaps when linking databases in a CA system</li> <li>- CA very far away from becoming a standard</li> </ul>
Cost Structure	<b>Decrease in Cost Structure</b> <ul style="list-style-type: none"> <li>- BDA (full population examination) makes audit assignments more cost-efficient and effective</li> <li>- Charging lower audit fees results in clients buying more advisory services: added value beyond traditional audit</li> </ul> <b>Increase in Cost Structure</b> <ul style="list-style-type: none"> <li>- Compatible software is unavoidably costly</li> <li>- Extra cost of hiring data scientists and BD analysts</li> </ul>

# Changes within Audit due to Data Analysis

Major Changes	Characteristics
Competencies of the Auditor	<ul style="list-style-type: none"><li>- Required IT knowledge, especially on data extraction and BDA</li><li>- Ability to recognize relevant patterns</li><li>- Critical thinking</li><li>- Education of future auditors need to be reformed</li></ul>
Standards	<ul style="list-style-type: none"><li>- Standards do not restrict the use of BD and BDA</li><li>- Standards should inform auditors about the possibilities of BD and BDA</li><li>- Need to give out more guidance</li><li>- Fundamental standards need adjustments</li><li>- A new specific standard on BDA should be in place</li><li>- Willingness of the auditor remains crucial</li></ul>

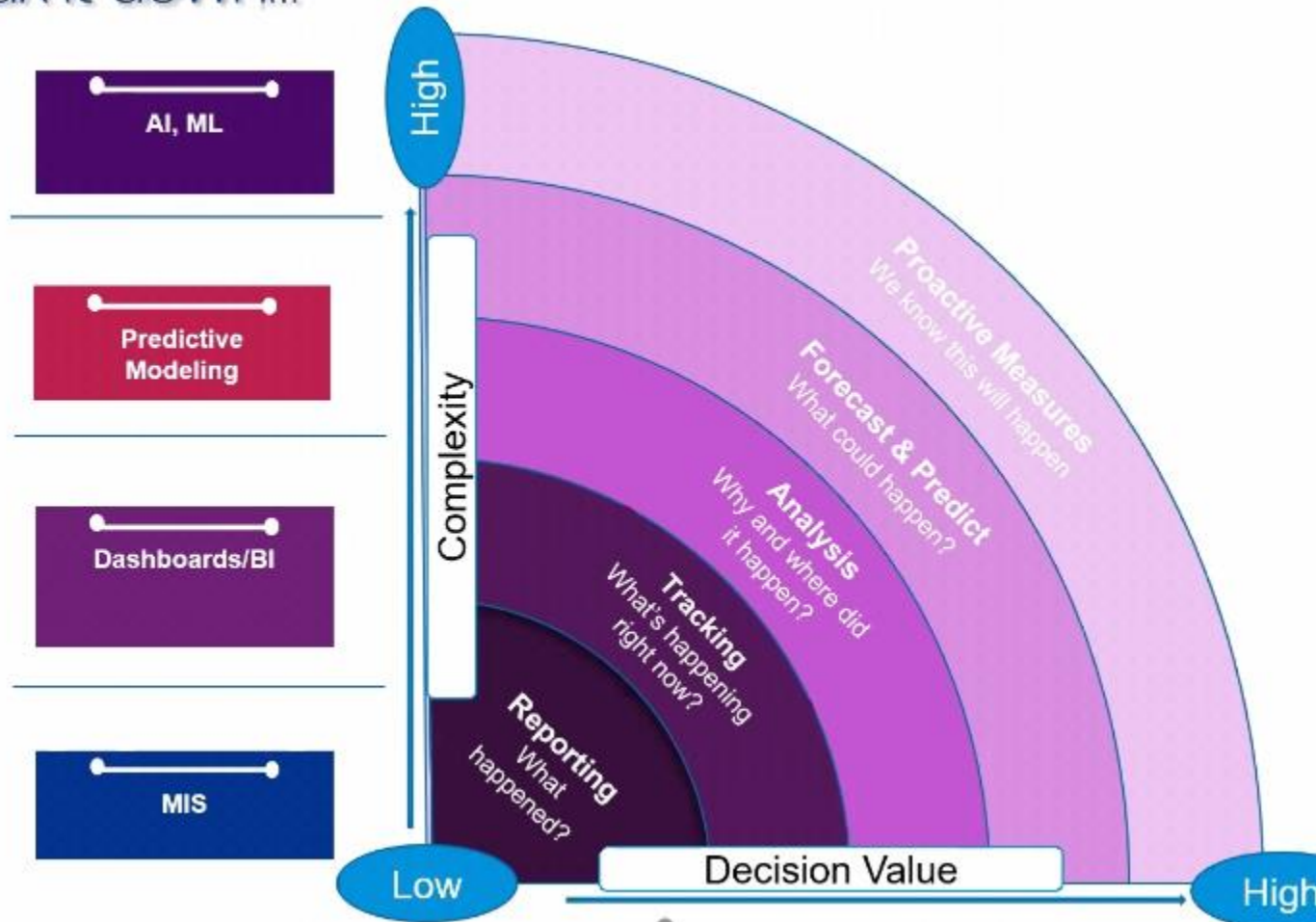


# Data Analysis for CXO's

Chief Risk Officer	Chief Tax Officer	Chief Information Officer	Chief Executive Officer	Chief Financial Officer
Anti-Money Laundering and KYC	Tax Forecasting	IT Ticket Analysis	Overall business performance analytics	Customer Data and Predictive Pricing
Fraud Detection/Prevention	Labelling Error detection in Retail	Software Optimization	Enhanced 360° Customer view	True Value of Assets
Streamlining Model Risk Management	Transaction detection	IT Inventory Management	Fraud Detection	Predicting and forecasting bad debts
Credit Risk and Loss Forecasting	Identifying tax deductions and credits	Unusual activity detection	Operations Analysis	Embezzlement and expense fraud
Targeted Risk Review		Log Analysis	Data Warehouse Augmentation	Money Laundering
			Customer Acquisition/Lead generation	Automating the Finance back-office
			Churn Reduction	

# New way for Analysis

Lets break it down...



# How we have progressed

## How have we progressed...

### Management Information Systems

Using computerized databases of information organized and programmed to produce regular reports to see:

*"What is happening?"*

#### Tools:

Decision Support Systems

Executive Information Systems

Marketing Information Systems

HR Management Systems

ERP Systems

### Descriptive Analytics

Using data aggregation and data mining to provide insight into the past and answer:

*"What has happened?"*

#### Tools:

++ ++ ++ ++ ++  
+ a b l e a u  
S O F T W A R E

Q alteryx

Power BI

### Artificial Intelligence

Using statistical models and forecasts techniques to understand the future and answer:

*"What could happen?"*

#### Tools:

R

python™

sas

TensorFlow

scikit-learn

K Keras

# Key Data analysis in Financial Statement

- V lookup
- Pivot Chart
- Ageing Analysis
- If and Or
- Conditional Formatting
- Ratio Analysis
- Expenses Grouping / GL Mapping
- Data Analysis
- Provision Mapping

# Key Ratio for stakeholders

1. **Gross margin percentage** =  $\text{Gross margin} / \text{sales}$  (How high the gross margin percentage needs to be depends on how a business is organized and the other costs it has to support)
2. **Net operating margin percentage** =  $\text{EBIT} / \text{sales}$  (Businesses with high net operating margin percentages are typically stronger than those with a low percentage. The higher the better!)
3. **Operating leverage** =  $\text{contribution} / \text{fixed costs}$  (A business that has an operating leverage of 1 is generating just enough revenue to pay for its fixed costs. This would mean that there is no return for the owners)
4. **Financial Leverage**:  $\text{total capital employed} / \text{shareholder's equity}$  (the higher a business's financial leverage, the riskier it is because there is more debt to be repaid.)
5. **Total leverage** =  $\text{operating leverage} \times \text{financial leverage}$  (represents the total risk that a company carries in its present business. Total leverage tells you the total effect a given change in the business should have on the equity owners)

# Key Ratio for stakeholders

6. **Debt-to-equity ratio** = total liabilities / total equity (the lower the debt-to-equity ratio, the more conservative the financial structure of the company. The more conservative the financial structure of a company, the less risk there is. Now, less risk isn't always what an investor is looking for, so you'll have to determine your own level of risk. This key ratio will help you know if a potential investment is meeting or exceeding that level of acceptable risk.)

7. Quick and current ratios

Quick ratio = liquid assets / current Liabilities

current ratio = current assets / current liabilities (If a company doesn't have enough current assets to cover its current liabilities, it is usually a sign of impending trouble. On the other hand, a current ratio and a quick ratio of 2 to 1 or higher is more appropriate)



# Financial Dashboard - Management

## FINANCIALS

## M ORDER BOOK STATUS

Previous Year

**INR 1,700**

Current Year

**TARGET**

**ACTUAL**

INR 2,000

INR 500

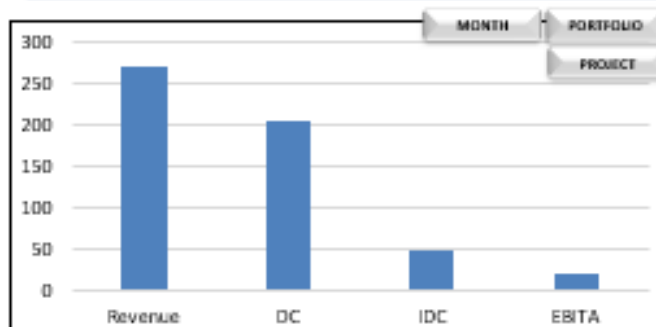
**Q** Number NIT released and not identified

3

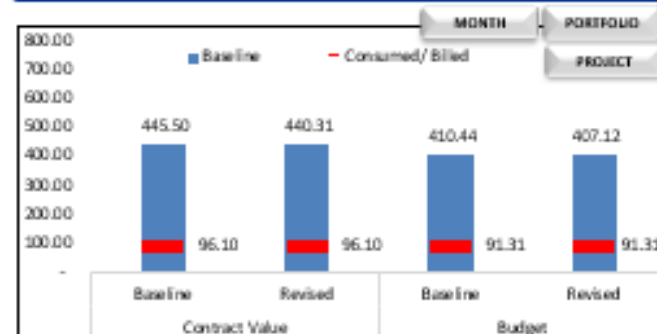
**Q** Number of wins with High Opportunity Score

2

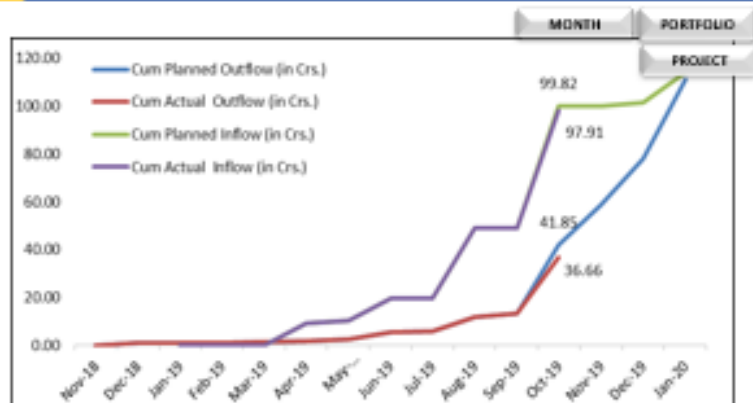
## PROFITABILITY AT PORTFOLIO & PROJECT LEVEL

**PROFIT & LOSS STATEMENT (In Crores)**

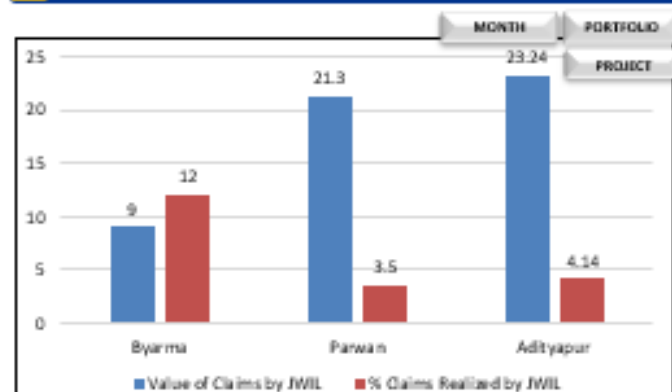
## PROJECT BUDGET COMPLIANCE



F	CASHFLOW (In Crores)
---	----------------------



## CLAIMS MANAGEMENT



### BONUS ACHIEVEMENT STATUS

Project Name	Contractual Period	Forecasted Completion	Potential Bonus (In Crores)	Bonus Achieved (In Crores)

Any Queries???