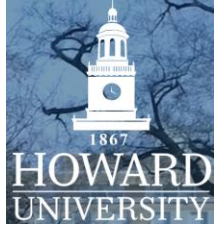


Virtual Applied Data Science Training Institute - Spring 2024 Training Series



Session 2:

Data Exploration and Visualization Using Power BI Session

Session Leader

Dr. Anthony Tsetse

(Associate Professor)

Northern Kentucky University, KY

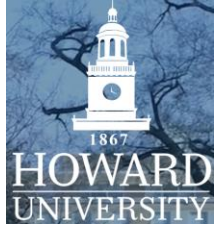
Instructor:

Dr. S. Tweneboah-Koduah

(Assistant Professor, Computer Science)

Gannon University, Erie, PA

Data Exploration and Visualization Using Power BI



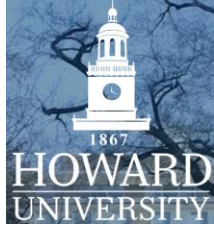
Day 1: Tuesday February 13

Outline

- Ω Session Overview and Expectations
- Ω Overview of Data Science
- Ω Short Break
- Ω Introducing Data Exploration
- Ω Open Access Data Repositories
- Ω Questions and Answers

Compiled by:
S. Tweneboah-Koduah, PhD
CIS, Gannon

Data Exploration and Visualization Using Power BI

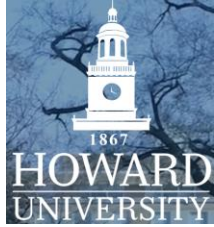


Day 2: Thursday February 15

Outline

- Ω Overview of Data Visualization
 - a. 1-Variable Graphs
 - b. 2-Variable Graphs
 - c. 3- and multivariable Graphs
- Ω An Overview of Power BI
 - a. The parts of Power BI
 - b. Use of Power BI and roles
 - c. Power BI flow
 - d. Use Power BI
 - e. Building blocks of Power BI
 - f. Power BI Services
 - g. Power BI license

Data Exploration and Visualization Using Power BI

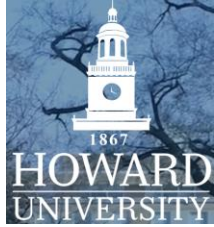


Day 3: Tuesday February 20

Outline

- Ω Importing Dataset and Modelling
 - a. Importing data into Power BI Desktop
 - b. Dealing with errors
 - c. 'Applied steps' in modelling data
- Ω Building Visuals and Dashboards Using Power BI
 - a. Plotting
 - i. Visualization Panel
 - ii. Plot modifications
 - iii. Pages
 - b. Explore Marketplace for other visualizations
 - i. Filters
 - ii. Edit interactions
 - c. Saving and Exporting 39
 - i. Saving as pbix files
 - ii. Exporting and publishing report

Data Exploration and Visualization Using Power BI



Day : Thursday February 22

Outline

- Ω Presentations (Participants)
- Ω Closing Session

Data Exploration and Visualization Using Power BI

Overview of Data Science/Big Data

Topics For Today's Session



Why Big Data?

Explosion of Data



- We create **2.5 quintillions** of data everyday.
- Data comes in from various sources and has different formats
- This enormous collection of data is **Big Data**.



Big Data Why the interest?

- One of the hottest topics in the world of business,,,,,,, for **business Intelligence**
 - Data is the foundation of any successful Business
 - Business analytics typically implies the analysis of very large data sets. (For this reason, the term **Data Science or Big Data**
- Today almost every business needs **Data Scientist**
- Almost every University of higher learning is running **Data Science** Program or related!!!



Google server farms... over 10 million of connected machines....locations across the globe



Data Science.....why the interest?

- Living in the age of technology has implications for everyone in the digital ecosystem.
 - Technology makes it possible to collect huge amounts of data.
 - Technology has given more people the power and responsibility to analyze data and make decisions.
- A large amount of data already exists and will only increase in the future.



Why Big Data?

(Data makes everything clearer)



- Breast cancer Disease
- Liver cancer Disease
- Prostate cancer Disease
- Lung cancer Disease
- Kidney cancer Disease

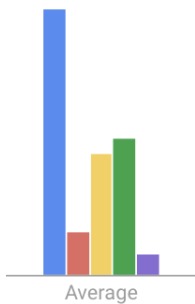
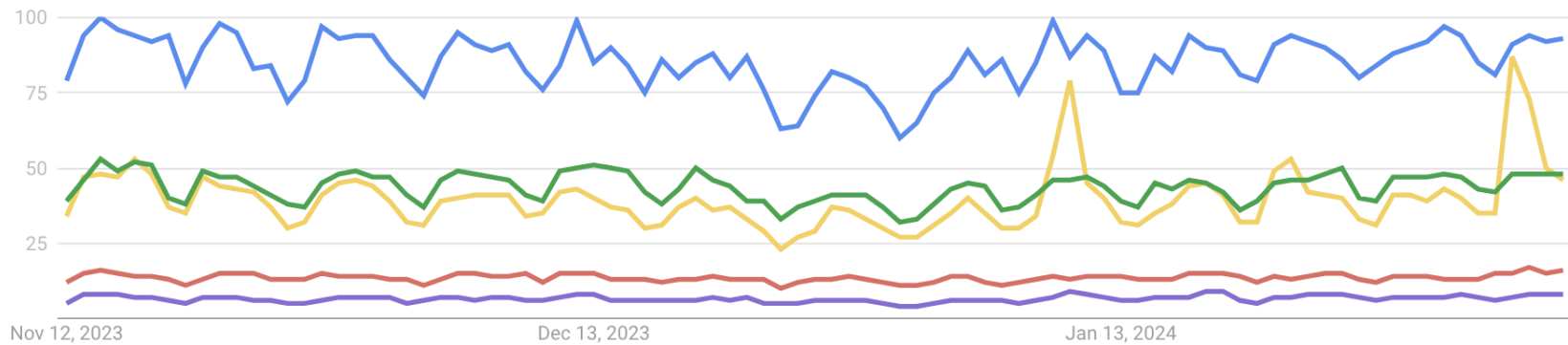
Worldwide ▼

Past 90 days ▼

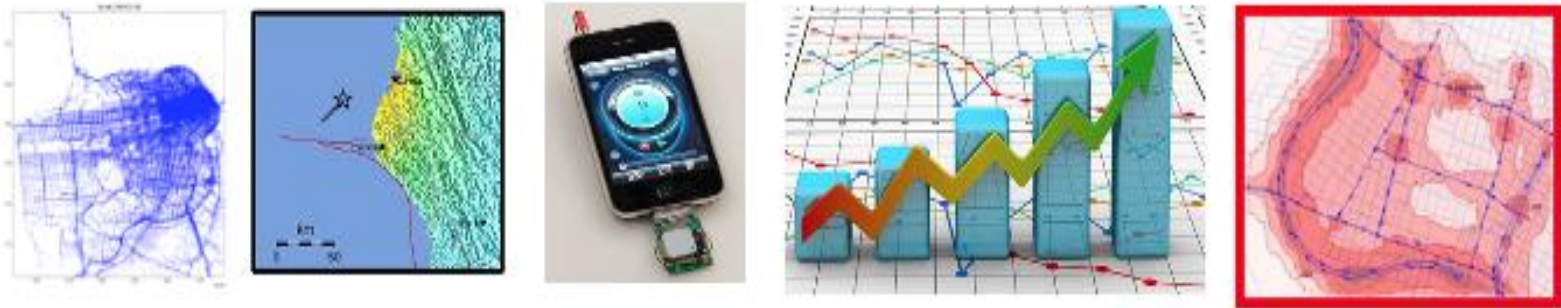
All categories ▼

Web Search ▼

Interest over time ?



“Big Data” the needs



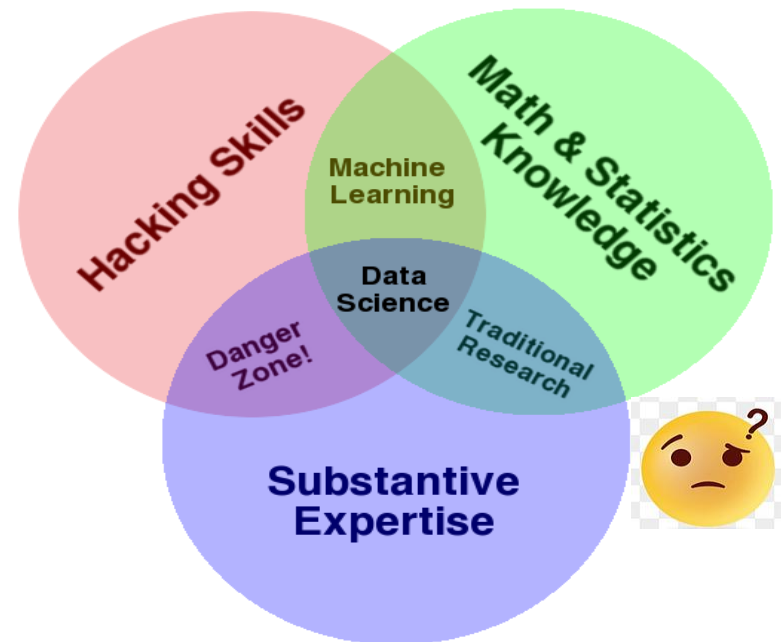
Crowdsourcing + physical modeling + sensing + forecasting + data assimilation +....



.....+..Planning

"Data Science/Big Data - What is it?"

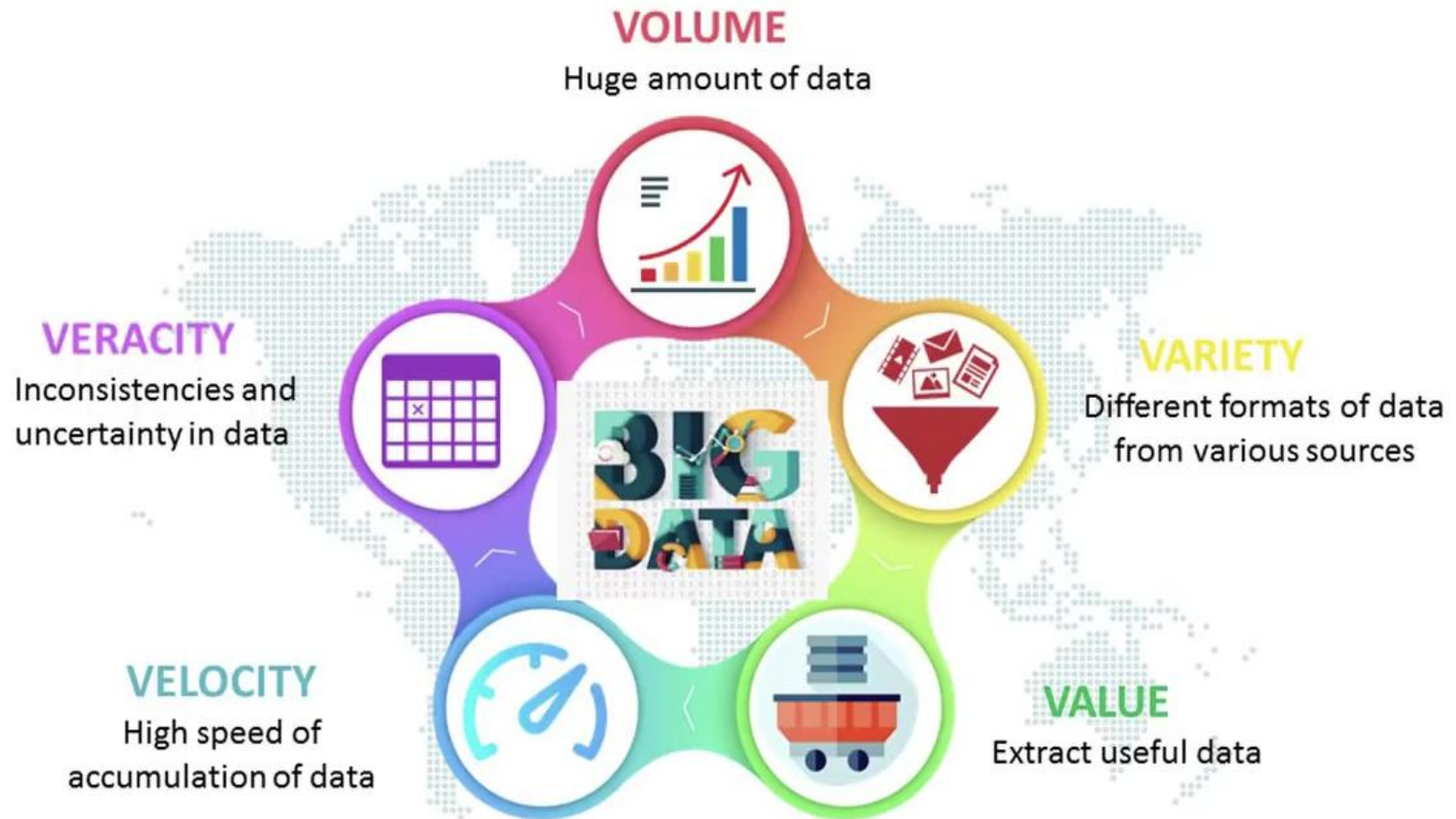
Multiple definitions.... new ones still emerging



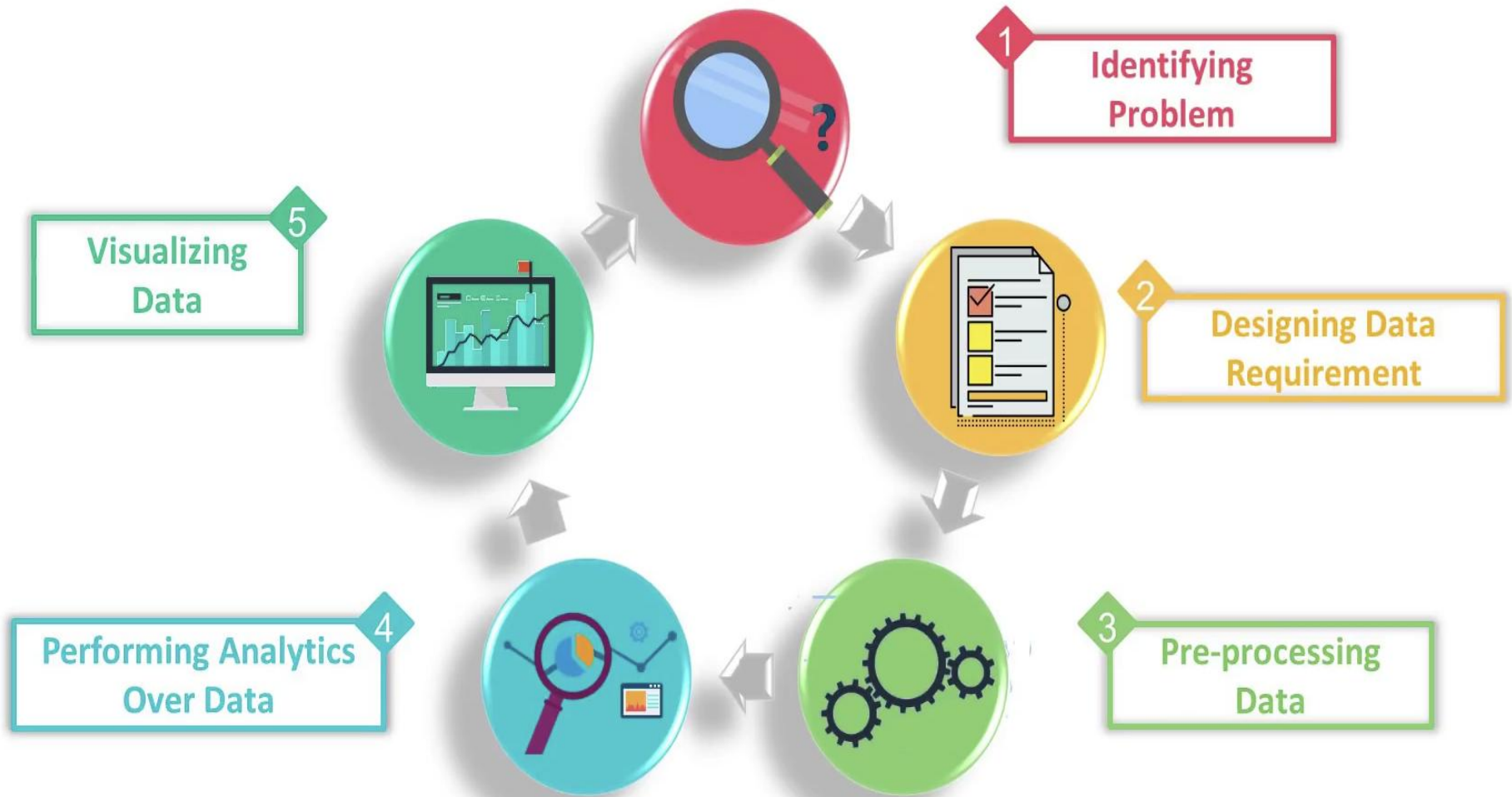
Data Science =

Mining Data (to gain insight and) **to make**
(informed) **Decisions** (knowledge)

Big Data – Key Characteristics



Data Analytics– Major Stages



“Big Data”: Sources

It's All Happening On-line

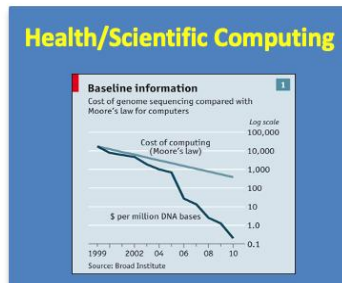


Every:
Click
Ad impression
Billing event
Fast Forward, pause,...
Server request
Transaction
Network message
Fault
...

User Generated (Web & Mobile)



Internet of Things / M2M



- Medical/Hospital Records
- Data marts/warehouses
- Transactional databases (OLAP)
- Spatial and temporal data
- Time-series data
- Stream data
- Text databases & WWW
- Business: Web, e-commerce, transactions, stocks, ...
- Science: Remote sensing, bioinformatics, scientific simulation, ...
- Society & Social media and everyone: news, digital cameras

Big Data: Confluence of Disciplines

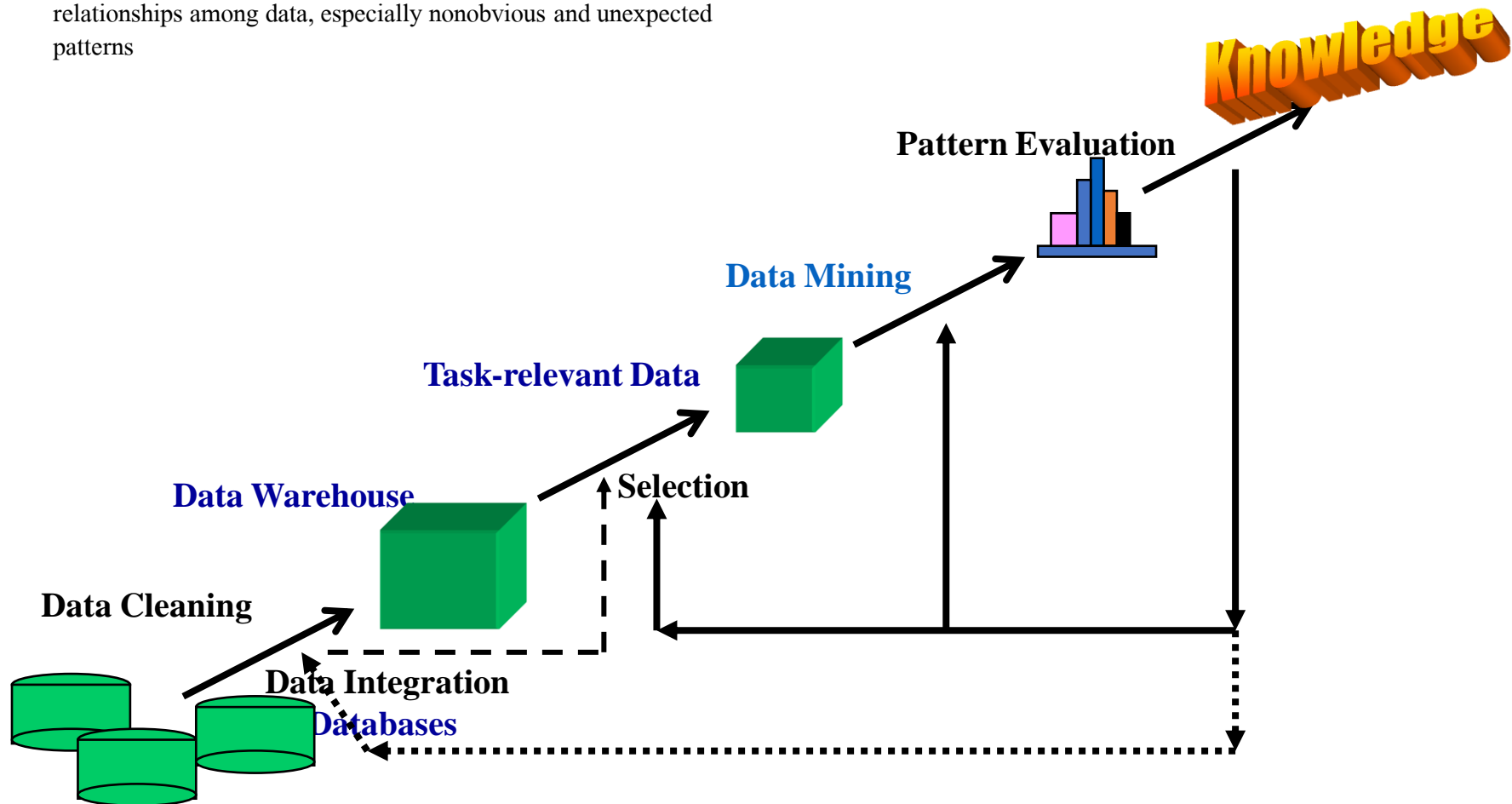


Data Science – Key Concepts

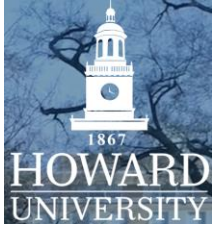
- Data Mining (Knowledge Discovery from Data)
- Techniques
- Tools
- Applications

Data Mining: Process

Data mining—attempts to discover patterns, trends, and relationships among data, especially nonobvious and unexpected patterns



Data Science – Methodologies!!!



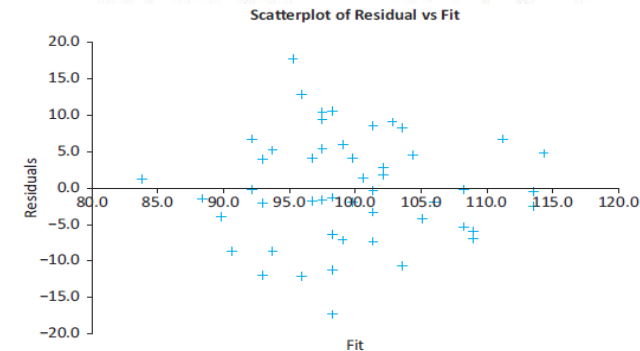
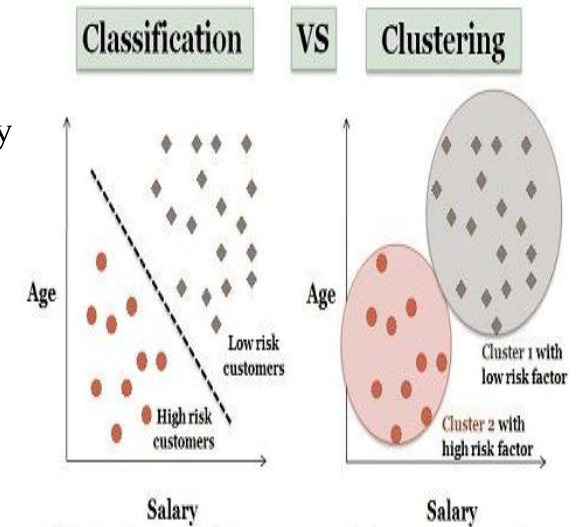
Once a data warehouse is in place, analysts can begin to mine the data with a collection of methodologies:

- *Classification analysis* attempts to find variables that are related to a categorical (often binary) variable.
- *Prediction* tries to find variables that help explain a continuous variable, rather than a categorical variable.
- *Cluster analysis* tries to group observations into clusters so that observations within a cluster are alike, and observations in different clusters are not alike.
- *Market basket analysis* tries to find products that customers purchase together in the same “market basket.”
- *Forecasting* used to predict values of a time series variable by extrapolating patterns seen in historical data into the future.

Data Science — Techniques

There are a number of techniques that can be applied in data science (usually for):
Modification, Storage, Analysis, Insights, and Representation (Visualization). They
include (not limited to.):

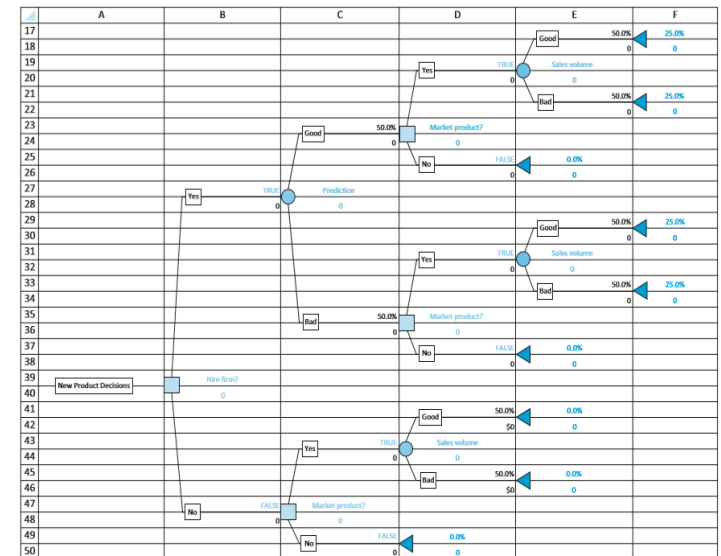
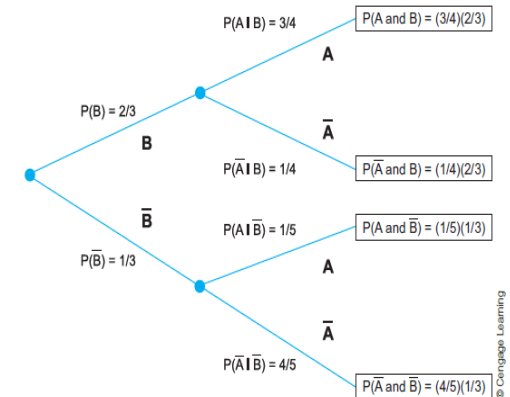
- Probability & Statistics (incl. Descriptive and Inferential statistics)
- Classification and Clustering (unsupervised learning)
- Anomaly Detection Analysis
- Regression (Linear & Multivariate) analysis



Graph Data	Sales	Fit	Residuals
1	85	83.8232949	1.176750604
2	103	108.9790334	-5.979033397
3	102	108.9790334	-6.979033397

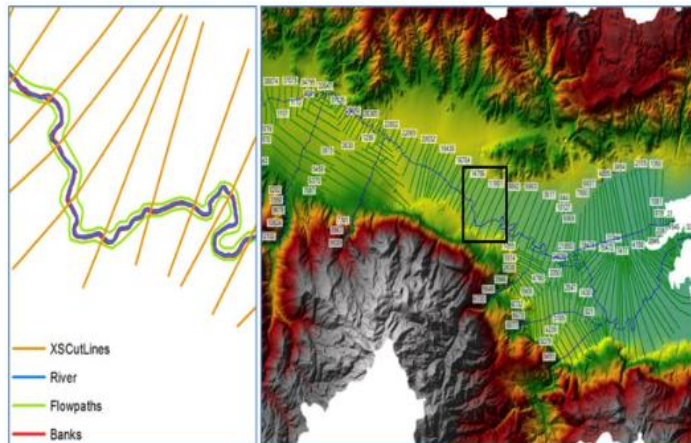
Data Science —Techniques (conti..)

- Regression Analysis
- Non-Parametric statistics
- Neural Networks
- K-Means clustering
- Probability and Decision Trees



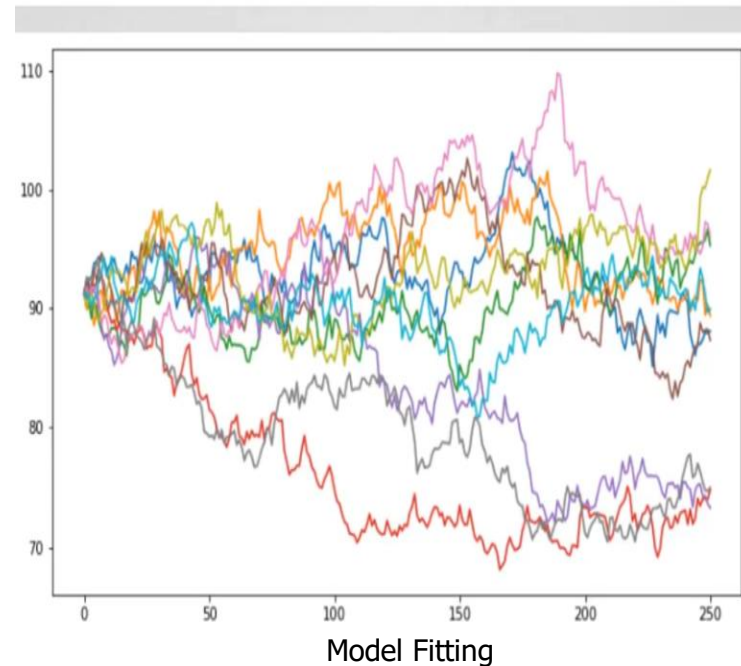
Data Science — Techniques (Conti..)

- Scala/Java
- Spatial Modelling
- Model Fitting



Spatial Modelling

$$S_t = S_{t-1} \cdot e^{((r - \frac{1}{2} \cdot \text{stdev}^2) \cdot \delta_t + \text{stdev} \cdot \sqrt{\delta_t} \cdot Z_t)}$$



Data Exploratory and Analytics Tools

■ Data Collection Tools

- Semantria, Trackur

■ Data Storage Tools

- Apache Hadoop, Apache Cassandra, Hbase, Mongo DB

■ Data Extraction Tools

- OctoParse, Content Grabber

■ Data Cleaning/Refining Tools

- OpenRefine

■ Data Analysis Tools

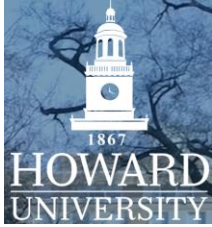
- Python, R, SAS, Excel Miner, Apache Spark

■ Data Visualization Tools

- **Power BI**, Python, Tableau, Google Fusion, Tableau



Modeling and Models



- A model is an abstraction of a real problem that tries to capture the essence and key features of the problem.
- Types of models; each can be a valuable aid in solving a particular problem:
 - Graphical models
 - Algebraic models
 - Spreadsheet models

Graphical Models (Visualization)

- Graphical models attempt to portray graphically how different elements of a problem are related—what affects what.
 - Do not provide enough quantitative details to “solve” the company’s problem
 - Purpose is usually to show the important elements of a problem and how they are related
 - Can be very enlightening for complex problems as information for management

Algebraic Models

- Algebraic models use algebraic equations and inequalities to specify a set of relationships in a very precise way.
 - A typical example is the “product mix” model shown below.

$$\max \sum_{j=1}^n p_j x_j$$

$$\text{subject to } \sum_{j=1}^n a_{ij} x_j \leq b_i \quad 1 \leq i \leq m$$

$$0 \leq x_j \leq u_j \quad 1 \leq j \leq n$$

Spreadsheet Models

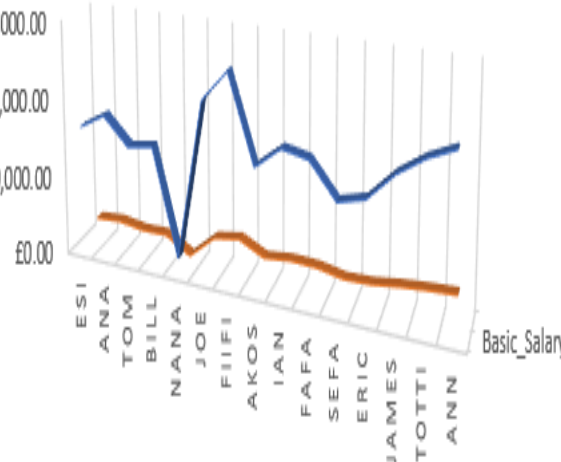
- Spreadsheet modeling is an alternative to algebraic modeling that relates various quantities in a spreadsheet with cell formulas.
 - Instant feedback is available from spreadsheets.
 - If a formula is entered incorrectly, it is often immediately obvious.
 - Developing good spreadsheet models is not easy.
 - They must be correct, well designed, and well documented.

Spreadsheet Models

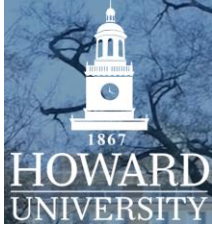
Names	Basic_Salary	Benefits	Total_Income	State_Tax	Local_Tax	Other_Ded	Total_Deductions	Net_Income	Tot_Annual_Basic	Tot_Ann_Benefits	Total_Ann_Income	Tot_Paid_Tax_State	Tot_Paid_Tax_Local	Tot_Ann_Deductions	Tot_Ann_net_Income
Esi	£82,005.00	£12,300.75	£94,305.75	£6,129.87						£147,609.00	£1,131,669.00	£73,558.49	£35,308.07	£231,602.85	£900,066.15
Ana	£92,000.00	£13,800.00	£105,800.00	£6,877.00						£165,600.00	£1,269,600.00	£82,524.00	£39,611.52	£259,831.26	£1,009,768.74
Tom	£75,000.00	£11,250.00	£86,250.00	£5,606.25						£135,000.00	£1,035,000.00	£67,275.00	£32,292.00	£211,818.96	£823,181.04
Bill	£78,000.00	£11,700.00	£89,700.00	£5,830.50						£140,400.00	£1,076,400.00	£69,966.00	£33,583.68	£220,291.72	£856,108.28
Nana	£10,500.00	£1,575.00	£12,075.00	£784.88						£18,900.00	£144,900.00	£9,418.50	£4,520.88	£29,654.65	£115,245.35
Joe	£112,567.00	£16,885.05	£129,452.05	£8,414.38						£202,620.60	£1,553,424.60	£100,972.60	£48,466.85	£317,917.66	£1,235,506.94
Fiifi	£132,900.00	£19,935.00	£152,835.00	£9,934.28						£239,220.00	£1,834,020.00	£119,211.30	£57,221.42	£375,343.20	£1,458,676.80
Akos	£78,000.00	£11,700.00	£89,700.00	£5,830.50						£140,400.00	£1,076,400.00	£69,966.00	£33,583.68	£220,291.72	£856,108.28
Ian	£92,123.00	£13,818.45	£105,941.45	£6,886.19						£165,821.40	£1,271,297.40	£82,634.33	£39,664.48	£260,178.64	£1,011,118.76
Fafa	£88,700.00	£13,305.00	£102,005.00	£6,630.33						£159,660.00	£1,224,060.00	£79,563.90	£38,190.67	£250,511.22	£973,548.78
Sefa	£67,000.00	£10,050.00	£77,050.00	£5,008.25						£120,600.00	£924,600.00	£60,099.00	£28,847.52	£189,224.94	£735,375.06
Eric	£72,000.00	£10,800.00	£82,800.00	£5,382.00						£129,600.00	£993,600.00	£64,584.00	£31,000.32	£203,346.20	£790,253.80
James	£89,900.00	£13,485.00	£103,385.00	£6,720.03						£161,820.00	£1,240,620.00	£80,640.30	£38,707.34	£253,900.33	£986,719.67
Totti	£102,000.00	£15,300.00	£117,300.00	£7,624.50						£183,600.00	£1,407,600.00	£91,494.00	£43,917.12	£288,073.79	£1,119,526.21
Ann	£110,234.00	£16,535.10	£126,769.10	£8,239.99						£198,421.20	£1,521,229.20	£98,879.90	£47,462.35	£311,328.68	£1,209,900.52

STAFF SALARIES

Basic_Salary Benefits

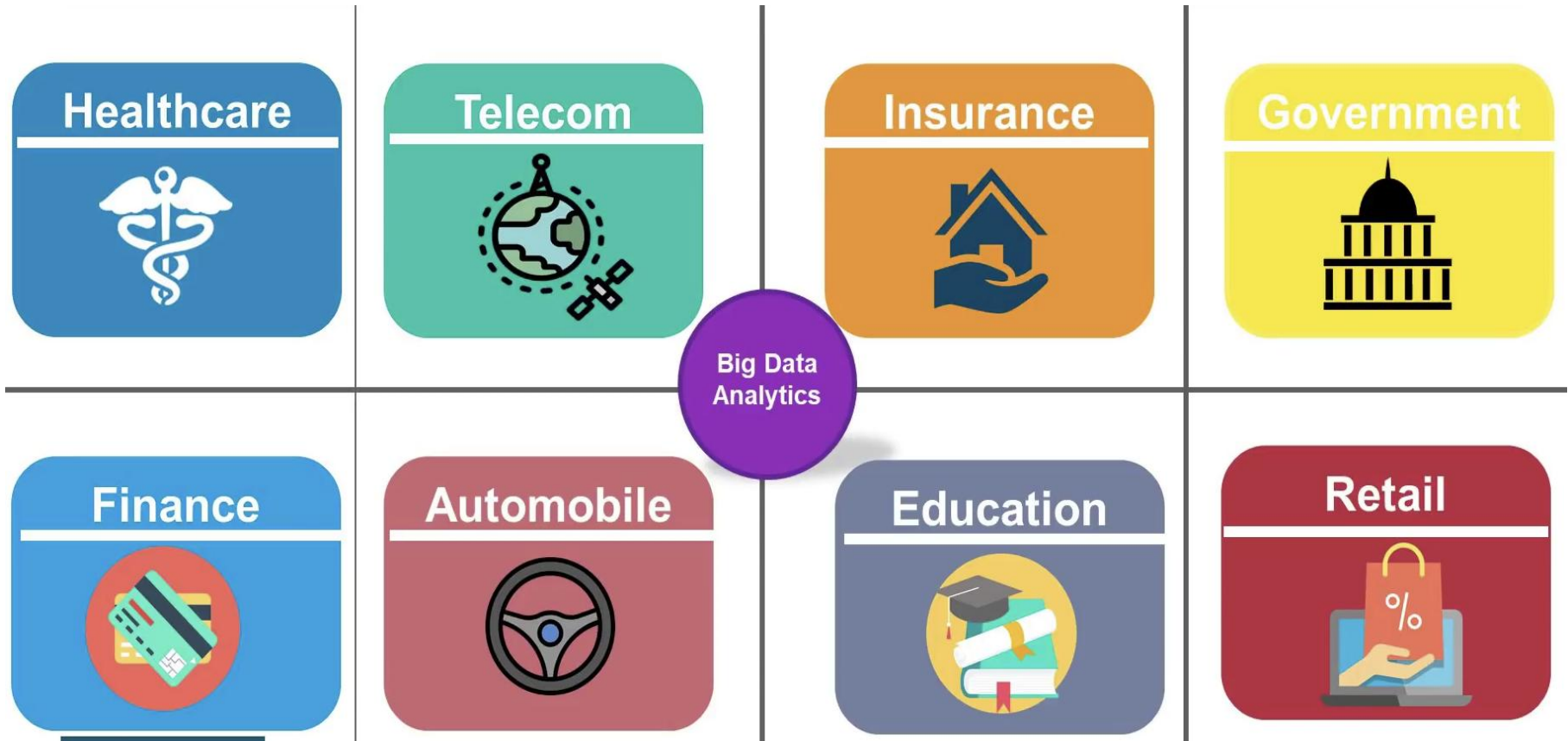


A Seven-Step DS Modeling Process

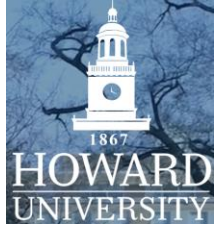


- Data Science portrays modeling as a seven-step process, but not all problems require all these seven steps.
 - Define the problem.
 - Collect and summarize data.
 - Develop a model.
 - Verify the model.
 - Select one or more suitable decisions.
 - Communicate your results.
 - Implement the model and update it over time.

Data Science: Applications



Data Science: Applications

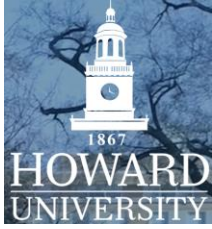


DS application areas are numerous, and new fields are emerging.....

▪ For Decision Support

- Market analysis
- Identifying and predicting disease
- Personalized healthcare recommendations
- Optimizing Shipping routes in real-time
- Customer relationship management (CRM), market basket analysis, market segmentation
- Search Optimization for Targeted Advertising and Re-targeting
- Price Comparison Website (e.g. pricecompare.com)

Market Analysis and Management



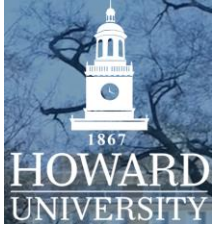
■ Where does the market data come from?

- Credit card transactions, discount coupons, customer complaint calls, phone calls log

■ Target marketing

- Find clusters of “model” customers who share the same characteristics: interest, income level, spending habits, etc.
- Determine customer purchasing patterns over time (e.g. Amazon, ebay, Walmart, Best Buy, etc)

Market Analysis and Management



■ Cross-market analysis

- Associations/co-relations between product sales, & prediction based on such association

■ Customer profiling

- What CUSTOMERS buy what PRODUCT?

■ Customer requirement analysis

- Identifying the best products for different customers
- Predict what factors will attract new customers

Risk, Fraud Detection & Mining Unusual Patterns

Approaches: Clustering & model construction for frauds, outlier analysis

Applications: Health care, retail, credit card service, telecom.

■ Medical insurance

- Patients, and ring of doctors
- Unnecessary or correlated screening tests

■ Telecommunications:

- Phone call model: destination of the call, duration, time of day or week.
- Analyze patterns that deviate from an expected norm

■ Retail industry

- Analysts estimate that 38% of retail shrink is due to dishonest employees and shoplifting

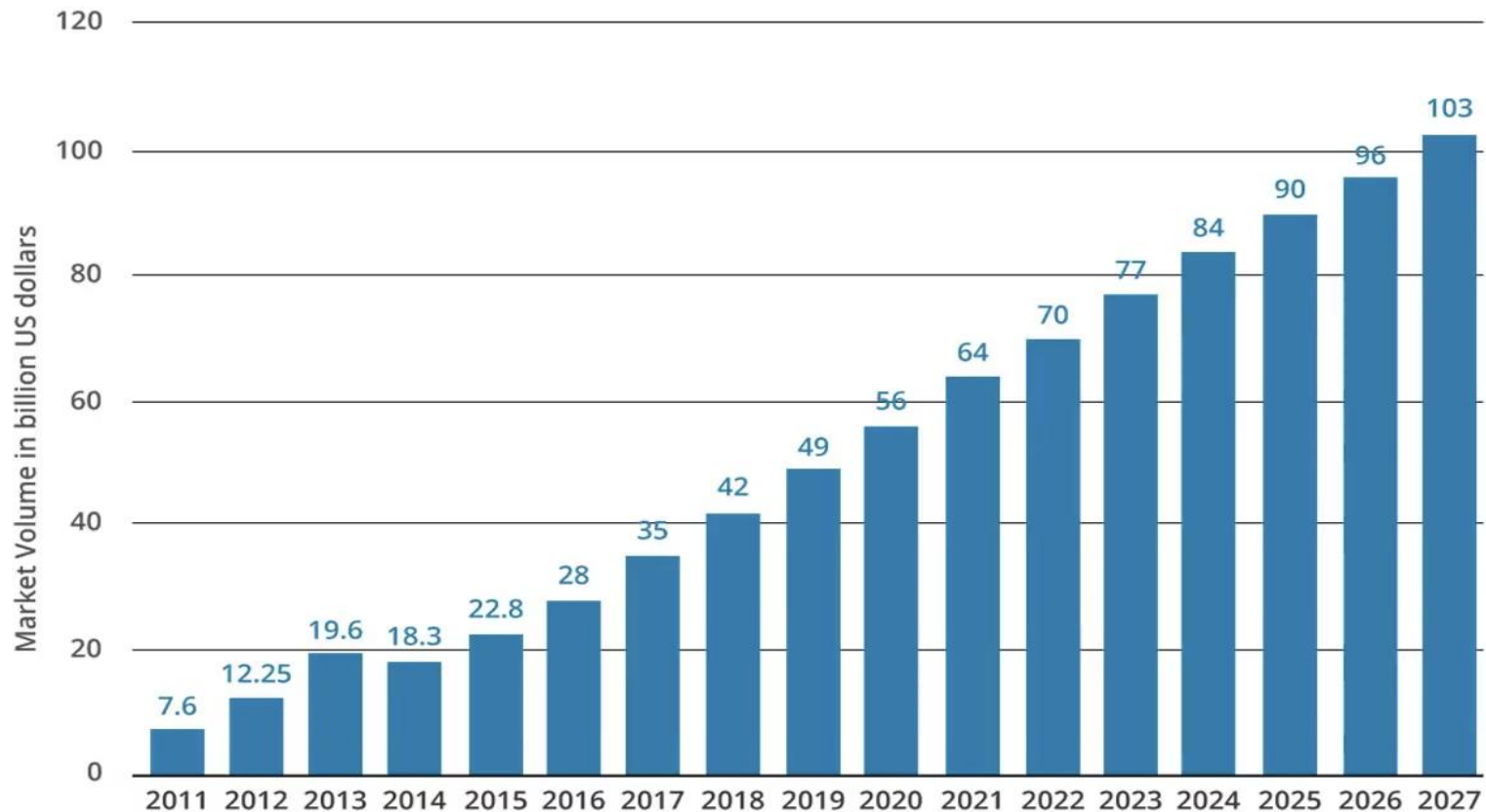
Data Mining – Applications (Conti..

Other Applications

- Text & Web mining (news group, email, documents)
- Data stream mining
- Bioinformatics and bio-data analysis
- Internet Search and Recommender Systems
- Image Recognition
- Speech Recognition
- Gaming/Gamification
- Airline Route Planning
- Delivery logistics

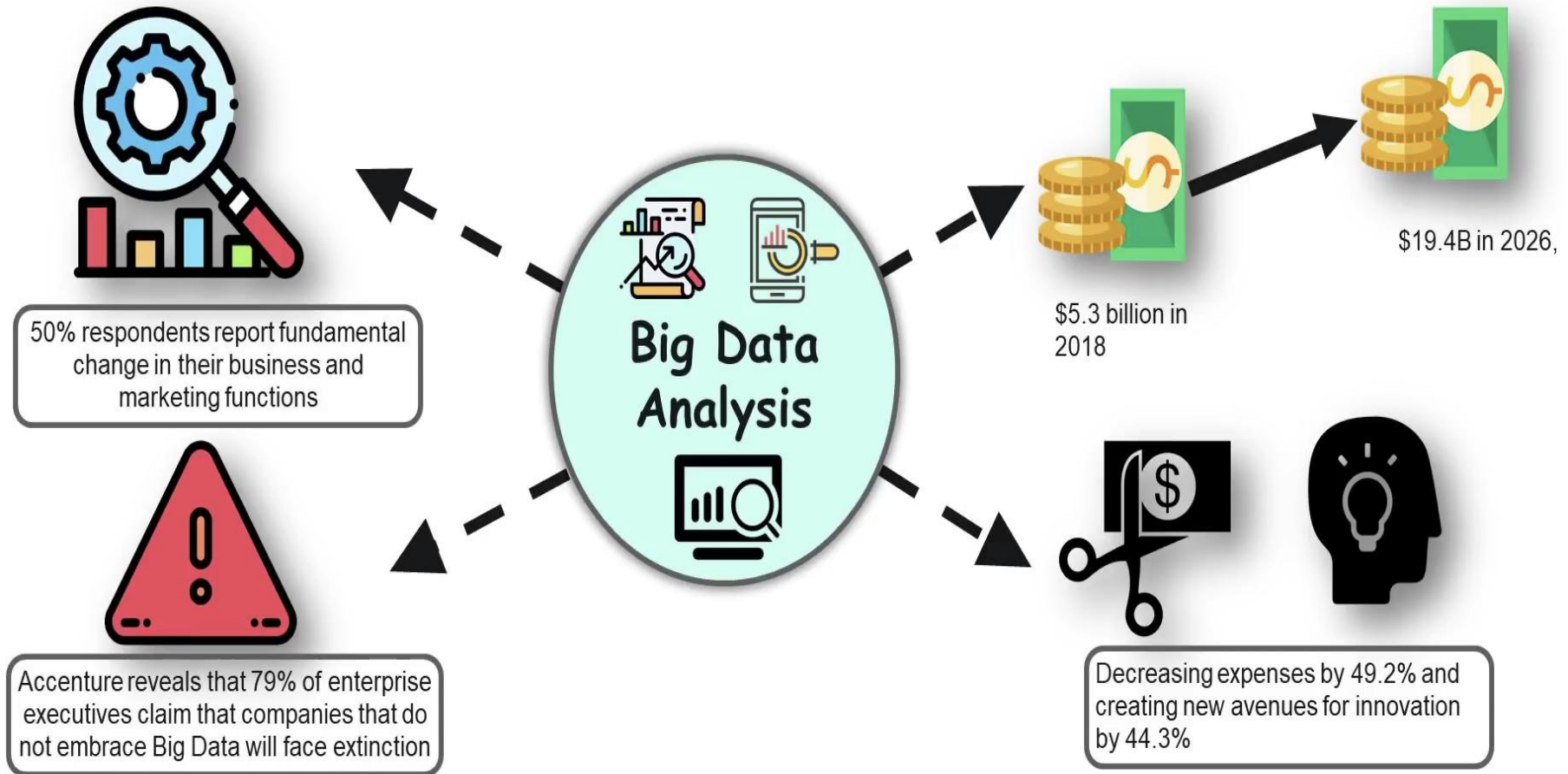
Trends in Data Science

- Big Data in Revenue Streams (According to Forbes)



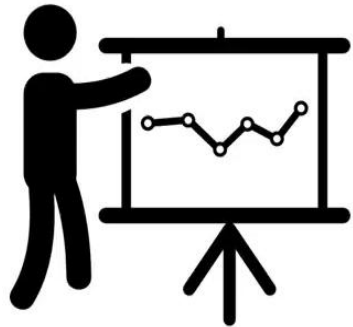
Trends in Data Science (conti...)

- Big Data in figures (According to Forbes)



Trends in Data Science (conti...)

• Big Data in figures (Career Opportunities)



Soaring Demand for Analytics Professionals



Salary Aspects



Huge Job Opportunities



Job Titles include :

- Big Data Analytics Business Consultant
- Big Data Analytics Architect
- Big Data Engineer
- Big Data Solution Architect
- Big Data Analyst
- Analytics Associate
- Business Intelligence and Analytics Consultant
- Metrics and Analytics Specialist

Trends in Data Science (conti...)

Required Skills Set (Data Analytics)

Here are a few skills which can be acquired depending upon the role in the field of Big Data Analytics :

Basic Programming



Statistical and quantitative Analysis



Data Warehousing

SQL

NoSQL



Data Visualization



Specific Business Knowledge



Computational Frameworks



The **TAKEAWAY**

- Your Comment, Contribution and Question

