



W E L C O M E



U.S. Department of Transportation
Office of the Assistant Secretary for
Research and Technology



Welcome



**Ken Leonard, Director
ITS Joint Program Office**
Ken.Leonard@dot.gov



www.pcb.its.dot.gov

Module 22:

Harnessing Social Media & Big Data Technologies for Transit Business Intelligence





Instructor



Susan Bregman

Principal

Oak Square Resources, LLC



Instructor



Manny Insignares
Vice President Technology
Consensus Systems Technologies



Learning Objectives

Define How Transit Providers Use Business Intelligence

Define Social Media Platforms and Their Applications to Public Transportation

Define Big Data in Relation to Social Media and Transit

Understand the Process for Applying Big Data Analytics to Social Media to Inform Transit Business Intelligence

Incorporate Findings to Support Business Intelligence with Data-driven Decisions



Learning Objective 1

Define how transit providers use business intelligence



Overview

- What is business intelligence?
- What are potential data sources?
- How can business intelligence benefit transit operators?



What is Business Intelligence?

- Combines information from multiple sources to support data-driven decisions
 - Quantitative – Ridership, fare revenue, mileage
 - Qualitative – Focus groups, interviews, social media
- Integrates data from internal and external sources
 - Internal – Automated vehicle location systems, customer panels
 - External – Social media posts, Census data
- Enables organizations to evaluate progress in achieving goals
- Supports internal decision-making



What are Potential Data Sources?

- Qualitative data (agency-generated)
 - Customer surveys and panels
 - Focus groups and stakeholder interviews
- Quantitative data (agency-generated)
 - Automatic passenger counting data (APC)
 - Automated vehicle location data (AVL)
 - General Transit Feed Specification files (GTFS and GTFS-rt)
 - Electronic fare payment system datasets (EFPS)
- External data sources
 - Social media posts
 - Census files and other public datasets



How Can Business Intelligence Benefit Transit Operators?

- Meet mandated reporting requirements
- Provide greater transparency in reporting to internal and external audiences
- Provide input for planning, operations, and capital investments
- Support briefings for senior staff and board of directors

How Can Business Intelligence Benefit Transit Operators?



EXAMPLE



GOAL

Improve customer satisfaction

ACTIONS

- Conduct surveys and focus groups.
- Establish online customer panel.
- Analyze social media posts to understand customer sentiment.

How Can Business Intelligence Benefit Transit Operators?



EXAMPLE

GOAL

Improve service reliability for bus operations

ACTIONS

- Review internal data on on-time performance and travel time.
- Examine social media posts to identify specific locations where bus routes are prone to delay.

How Can Business Intelligence Benefit Transit Operators?



EXAMPLE

GOAL

Improve maintenance at rail stations

ACTIONS

- Review internal maintenance records.
- Analyze social media posts to identify issues on specific vehicles or stations.
- Encourage customers to report issues via social media (e.g., broken lights, overflowing trash, disabled ticketing machines).



How Can Business Intelligence Benefit Transit Operators?



EXAMPLE

GOAL

Improve transparency in performance reporting.

ACTIONS

- Develop key performance indicators (KPI) from available data sources.
- Report KPIs via online performance dashboard.

ACTIVITY



Question

Which of the following is NOT a source of data for business intelligence?

Answer Choices

- a) Automatic passenger counters (APC)
- b) Social media posts
- c) Electronic fare collection systems (EFCS)
- d) None of the above

Review of Answers



a) Automatic passenger counters (APC)

Incorrect. APC data can be analyzed to support agency decision-making.



b) Social media posts

Incorrect. Social media posts can be analyzed to support agency decision-making.



c) Electronic fare collection systems data (EFCS)

Incorrect. EFCS data can be analyzed to support agency decision-making.



d) None of the above

Correct! All the data sources listed can be used to support transit decision-making.



Learning Objective 2

Define social media platforms and their applications to public transportation



Overview

- What is social media?
- Taxonomy of social media platforms
- Use of social media by transit operators for agency-generated information
- Use of social media by transit customers and stakeholders for user-generated information
- Use of crowdsourcing and peer-to-peer platforms for sharing communication about transit

What is Social Media?



- Social media platforms are web-based or mobile applications that encourage users to interact with (and often influence) one another in real time.
- Social media, also called social networking, includes different types of applications.
- Platforms are mostly owned by private companies with proprietary formats and are not consistently regulated.
- Social media posts can share information (and misinformation).
- Social media is still evolving, and platforms continue to change.

Taxonomy of Social Media

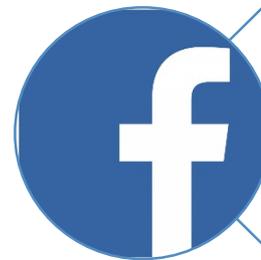
- Social networks
- Media sharing networks
- Discussion forums
- Content curation
- Consumer review networks
- Blogging and publishing networks



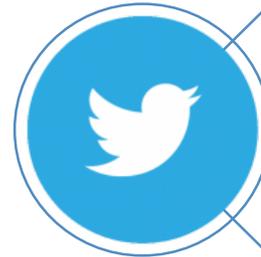
Taxonomy of Social Media

Social Networks

- Connect with other people online
- Share information, comments, and media
- Personal and professional networks



Facebook



Twitter

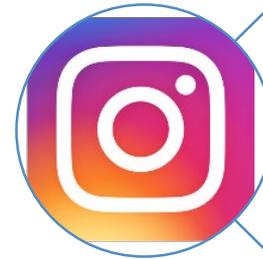


LinkedIn

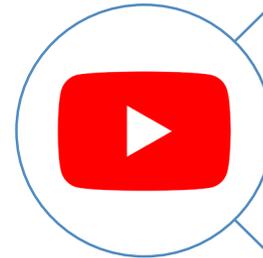
Taxonomy of Social Media

Media-Sharing Networks

- Share images, videos, and other types of media with others.
- Offer comments and other forms of feedback.



Instagram



YouTube



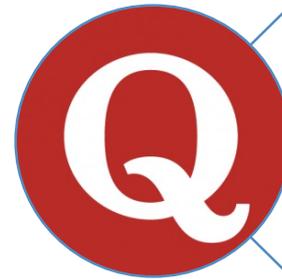
Vimeo

Discussion Forums

- Platforms serve as discussion boards
- Users can ask and answer questions, share information, and participate in discussions



reddit



Quora

Content Curation Platforms

- Identify and share content from multiple sources
- Content types include photographs, graphics, videos, presentations, and text



Pinterest

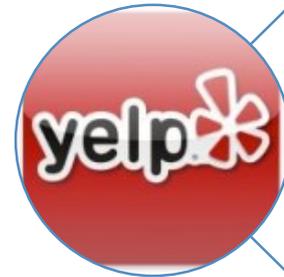


SlideShare

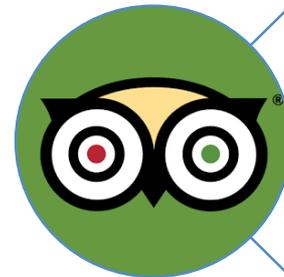
Taxonomy of Social Media

Consumer Review Networks

- Generate reviews and share opinions about goods and services.
- Most consumer websites also include customer reviews (e.g., Amazon).



Yelp

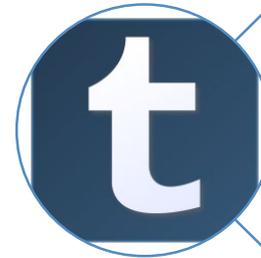


TripAdvisor

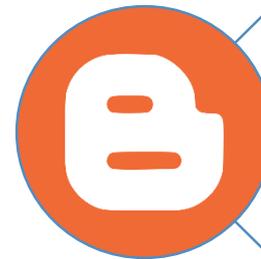
Taxonomy of Social Media

Blogging and Publishing Networks

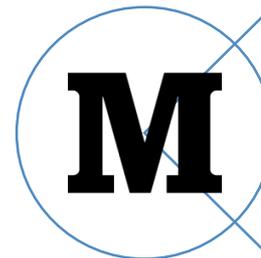
- Create content on user-defined topics.
- Posts are typically longer than most social networking sites.
- Organizations may use platforms to share news.



Tumblr



Blogger



Medium



Agency-Generated Social Media

Overview

- Most transit operators use social media for **outbound** communications.
 - Service updates and alerts
 - Emergency communications
 - Marketing activities
 - Customer service
 - Solicit customer feedback
 - General agency communications
- Audiences may include riders, stakeholders, media, first responders, public officials, and community members.



Outbound communications typically do not support business intelligence activities.



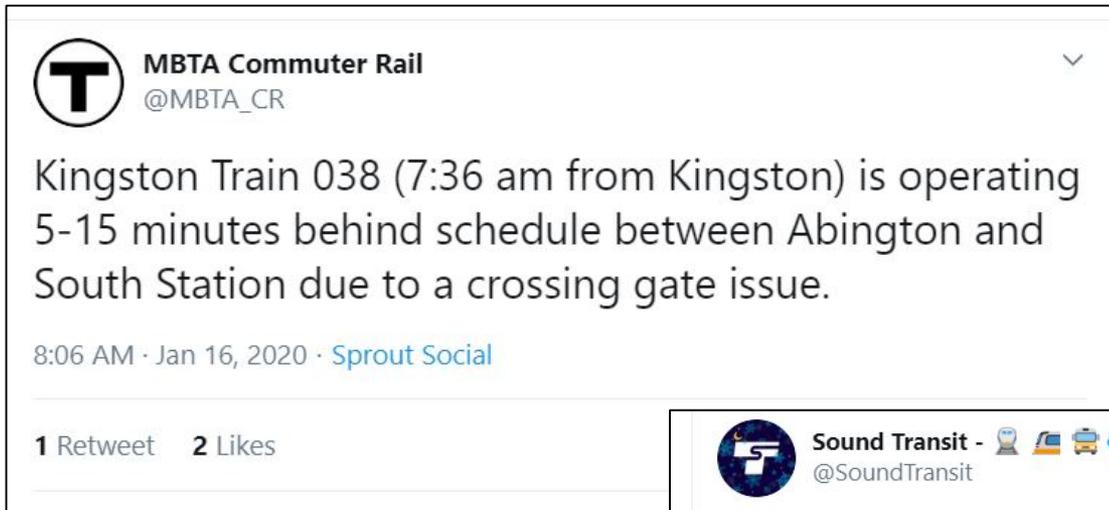
Agency-Generated Social Media

Service Updates and Alerts

- Notify customers about service changes
 - Planned
 - Unplanned
- Provide information about traffic delays and construction impacts
- Provide details about service during special events
- Twitter is especially well-suited for real-time alerts

Agency-Generated Social Media

Service Updates and Alerts



 **MBTA Commuter Rail**
@MBTA_CR

Kingston Train 038 (7:36 am from Kingston) is operating 5-15 minutes behind schedule between Abington and South Station due to a crossing gate issue.

8:06 AM · Jan 16, 2020 · [Sprout Social](#)

1 Retweet 2 Likes



 **Sound Transit -** 
@SoundTransit

Elevator Alert - The Pioneer Square Station south end mezzanine to surface elevator is out of service

 Elevator Alert - The Pioneer Square Station south end mezzanine to surface elevator is out of...
content.govdelivery.com

10:23 AM · Jan 16, 2020 · [GovDelivery](#)



Agency-Generated Social Media

Emergency Communications

- Use social media to communicate during health emergencies, weather events, and natural disasters (e.g., COVID-19, hurricanes, earthquakes).
- Use social media to share public safety information (e.g., Amber alerts, criminal activity).
- Twitter is especially well-suited for real-time alerts.

Agency-Generated Social Media

COVID-19 Pandemic Communications

Pinned Tweet

 **NYCT Subway. Stay Home. Stop the Spread.** ✓
@NYCTSubway

We're operating subway service for essential trips only.
If you must travel, be sure to:

- 1 Check [MTA.info](https://www.mta.info) or MYmta to see how often your line is running.
- 2 Wear a face covering.
- 3 Allow extra time to get where you're going.

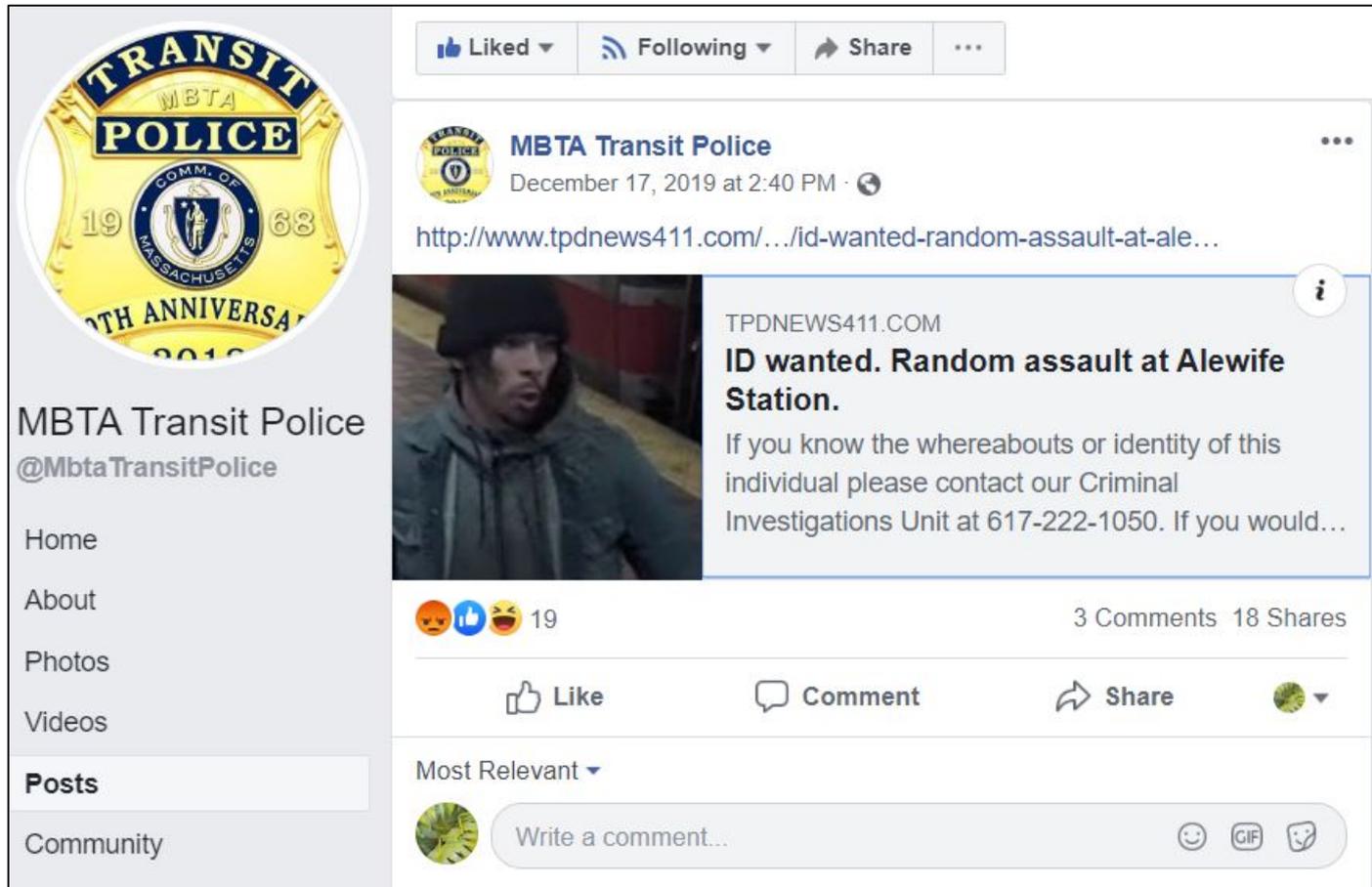
If you need help, @ or DM us. 24/7.



10:57 PM · Apr 27, 2020 · [Twitter Web App](#)

Agency-Generated Social Media

Public Safety Communications



MBTA Transit Police
@MbtaTransitPolice

Home
About
Photos
Videos
Posts
Community

Liked Following Share ...

MBTA Transit Police
December 17, 2019 at 2:40 PM · 🌐

<http://www.tpdnews411.com/.../id-wanted-random-assault-at-ale...>

TPDNEWS411.COM
ID wanted. Random assault at Alewife Station.
If you know the whereabouts or identity of this individual please contact our Criminal Investigations Unit at 617-222-1050. If you would...

19 reactions (3 angry, 1 like, 5 laughing) · 3 Comments · 18 Shares

Like Comment Share

Most Relevant

Write a comment... 🗨️ 🧠 🗣️



Agency-Generated Social Media

Marketing Activities

- Social media can help agencies create an image or identity.
- Media-sharing and blogging platforms are a good match for these posts.

Agency-Generated Social Media

Marketing Activities



M metrolosangeles • Following

M metrolosangeles #tbt: many years ago, RTD bus operators trained in the LA River. #transithistory #GoMetro

6d

subculture891 !!!!!##No way!!!!#
6d 4 likes Reply

jonpwong For a second, I thought it was a behind the scenes shot of the movie Speed
6d 12 likes Reply
View replies (2)

fit_cooki3 I just got trained the

♡ 💬 📍 📌

Liked by nickleonardphoto and 3,724 others

6 DAYS AGO

Add a comment...

Post



Agency-Generated Social Media

Customer Service

- Provide real-time customer service.
- Address customer comments and complaints.

Agency-Generated Social Media

Customer Service



Erin Rush @eerrush · 21m
A little wind and @SEPTA just falls apart... My 45 minute commute home should not be almost 3 hours 😞

1 reply Retweet Like Share

SEPTA_SOCIAL ✓
@SEPTA_SOCIAL

Replying to @eerrush

Sorry to hear this, Erin. Were you riding on the Paoli/Thorndale Line this evening? A downed tree caused issues on the line. Our apologies for the inconvenience. ^KG

6:26 PM · Jan 16, 2020 · [Hootsuite Inc.](#)



Agency-Generated Social Media

Solicit Customer Feedback

- Use social media to reach out to customers.
- Seek feedback on projects or programs.

Agency-Generated Social Media

Solicit Customer Feedback



**Talk Transit
with us.**

Let us know how to make the
UCLA/Westwood Commuter Express
service better for you.



longbeachtransit • Following
Skylinks at Long Beach

longbeachtransit Long Beach
Transit is holding a public meeting to listen to your valuable feedback about our UCLA/Westwood Commuter Express.

-
- Join us Wednesday, January 29 at 6:30pm at Skylinks
📍 4800 E Wardlow Rd, Long Beach, CA 90808. 🚌
-

If you can't make the meeting but would still like to give feedback, please email comments@lbt.com.

1d

52 likes
1 DAY AGO

Add a comment... Post



Agency-Generated Social Media

General Agency Announcements

- Share agency information
- Job listings
- Press releases
- Social posts can complement – but should not replace – traditional communications channels.

Agency-Generated Social Media

General Agency Announcements

Pay your fare
and watch cat
videos with the
same device.

OMNY
Just tap and go.
omny.info

DART

Dallas Area Rapid Transit (Official DART page)
@DARTDallas

- Home
- Photos
- Videos
- Posts
- Events
- About
- Community
- Get Email Updates
- Poll

Like Follow Share ...

Dallas Area Rapid Transit (Official DART page)
January 8 at 5:39 PM · 🌐

In 2019, Dallas Area Rapid Transit (DART) was far more than just the thing you ride. With 13 service area cities covering a 700 sq. mile area made up of 2.6 million citizens, more than 140 bus/shuttle routes, 11,000 bus stops, 14 On-Demand GoLink zones, 93 miles of light rail transit, 64 light rail stations, 5 commuter rail stations, and paratransit service for persons who are mobility impaired, DART continued to expand its mission to be your preferred choice of transportation for now and in the future.

DART.ORG
DART.org - DART News Release
Find more information about the agency in DART news releases.

👍 24 6 Comments 3 Shares



Customer-Generated Social Media

Overview

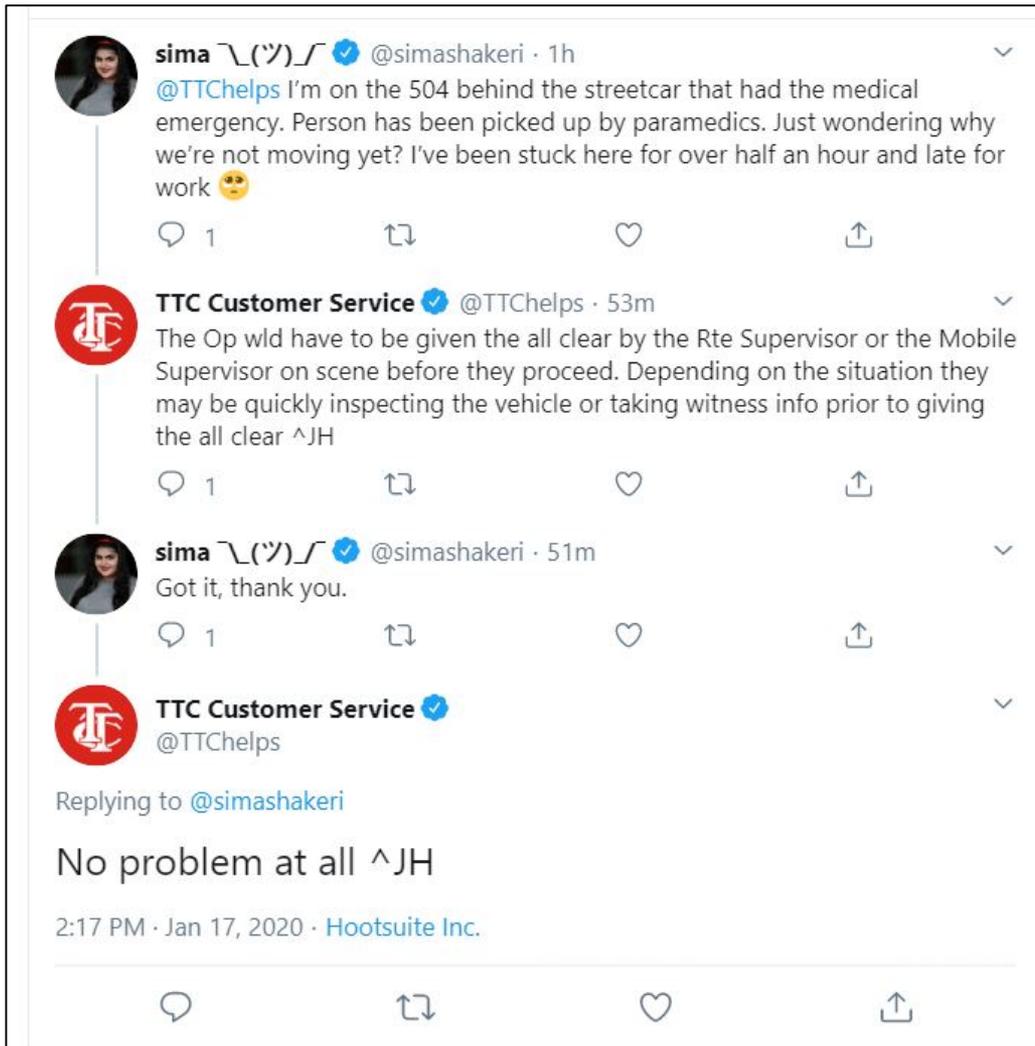
- Social media posts from transit customers, stakeholders, and others can provide unfiltered feedback
- User-generated posts typically include the following
 - Questions (e.g., where is the bus? what is the fare?)
 - Complaints (e.g., service, maintenance, safety, security)
 - Compliments (e.g., operator commendations)
- These inbound communications can be generated by riders, stakeholders, and community members and shared widely.



Organizations can use data mining techniques to analyze user-generated social media posts to support business intelligence activities.

Customer-Generated Social Media

Customer Questions



The screenshot shows a Twitter thread with four tweets. The first tweet is from user **sima** (@simashakeri) asking a question about a streetcar delay. The second tweet is a reply from **TTC Customer Service** (@TTChelps) explaining the procedure. The third tweet is a reply from **sima** saying "Got it, thank you." The fourth tweet is a reply from **TTC Customer Service** saying "No problem at all".

sima _(ツ)_/ ✓ @simashakeri · 1h
@TTChelps I'm on the 504 behind the streetcar that had the medical emergency. Person has been picked up by paramedics. Just wondering why we're not moving yet? I've been stuck here for over half an hour and late for work 😞

1 1 1 1

TTC Customer Service ✓ @TTChelps · 53m
The Op wld have to be given the all clear by the Rte Supervisor or the Mobile Supervisor on scene before they proceed. Depending on the situation they may be quickly inspecting the vehicle or taking witness info prior to giving the all clear ^JH

1 1 1 1

sima _(ツ)_/ ✓ @simashakeri · 51m
Got it, thank you.

1 1 1 1

TTC Customer Service ✓ @TTChelps
Replying to @simashakeri
No problem at all ^JH
2:17 PM · Jan 17, 2020 · Hootsuite Inc.

1 1 1 1

Customer-Generated Social Media

Customer Complaints

The screenshot shows a Twitter thread with four tweets. The first tweet is from 'mohamed atta boy' (@rosswaldz) asking why the Park Street escalator is broken. The second tweet is from 'MBTA' (@MBTA) asking for more details. The third tweet is from 'mohamed atta boy' specifying the escalator. The fourth tweet is from 'MBTA' replying with an apology and a promise to follow up.

mohamed atta boy 🌹🍞 @rosswaldz · 2h
@MBTA why is the park street escalator never operable for more than like 4 consecutive days at a time??

MBTA @MBTA · 2h
Hi and thanks for reaching out. Can you tell us which escalator you are referring to?

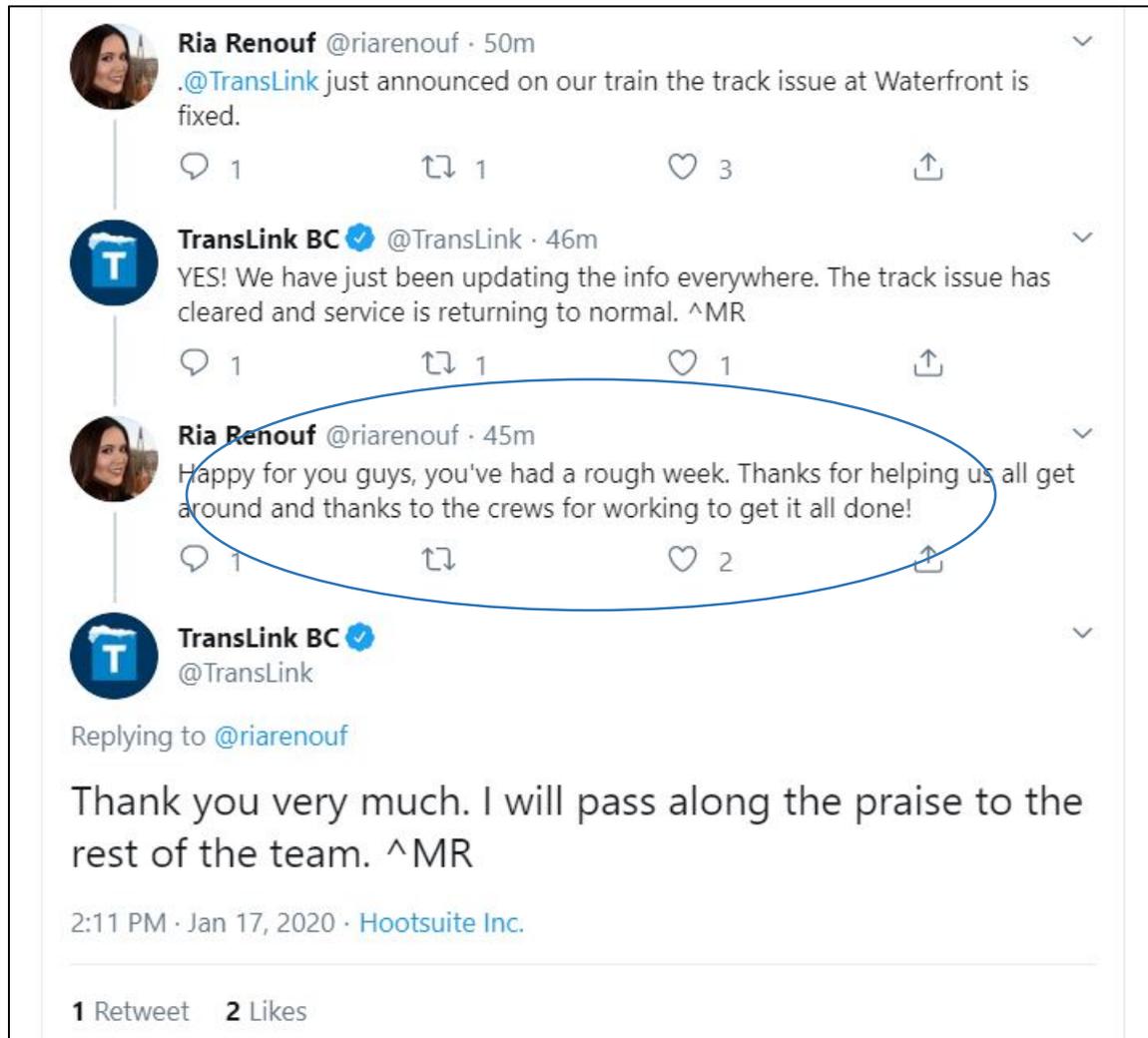
mohamed atta boy 🌹🍞 @rosswaldz · 2h
The one used to get out of the station from the eastbound green line arrivals

MBTA @MBTA
Replying to @rosswaldz
Thank you. This escalator was being worked on earlier and should be back in service. We'll follow-up with Station Maintenance.

1:00 PM · Jan 17, 2020 · TweetDeck

Customer-Generated Social Media

Customer Compliments



The screenshot displays a Twitter thread with four tweets. The third tweet, from Ria Renouf (@riarenouf), is circled in blue. The fourth tweet is a reply from TransLink BC (@TransLink) to Ria Renouf. The interface includes profile pictures, usernames, timestamps, text content, and engagement icons (reply, retweet, like, share).

Ria Renouf @riarenouf · 50m
.@TransLink just announced on our train the track issue at Waterfront is fixed.

1 Reply 1 Retweet 3 Likes

TransLink BC @TransLink · 46m
YES! We have just been updating the info everywhere. The track issue has cleared and service is returning to normal. ^MR

1 Reply 1 Retweet 1 Like

Ria Renouf @riarenouf · 45m
Happy for you guys, you've had a rough week. Thanks for helping us all get around and thanks to the crews for working to get it all done!

1 Reply 2 Likes

TransLink BC @TransLink
Replying to @riarenouf

Thank you very much. I will pass along the praise to the rest of the team. ^MR

2:11 PM · Jan 17, 2020 · [Hootsuite Inc.](#)

1 Retweet 2 Likes



Crowdsourcing and Peer-to-Peer Communications

Overview

- Crowdsourcing solicits ideas and feedback on a specific topic from a large group of people via the Internet.
- Some mobile applications create a platform for subscribers to share information with one another.

Crowdsourcing and Peer-to-Peer Communications

Overview

- Crowdsourcing solicits ideas and feedback on a specific topic from a large group of people via the Internet.
- Some mobile applications create a platform for subscribers to share information with one another.
- Examples include:
 - Transit – Mobile app complements real-time data feeds with crowdsourced info
 - Pigeon – Google app for crowdsourced info
 - Clever Commute – Mobile app for sharing customer info for NJ Transit, LIRR, MNR services



ACTIVITY



Question

Which of these is NOT a source of social media data for business intelligence?

Answer Choices

- a) Agency marketing posts
- b) Customer complaints
- c) Customer questions
- d) Peer-to-peer communications

Review of Answers



a) Agency marketing posts

Correct! Marketing social media posts can generate goodwill for an agency, but they are not used to inform data-driven decisions.



b) Customer complaints

Incorrect. Customer complaints can provide valuable data.



c) Customer questions

Incorrect. Customer questions can provide valuable data.



d) Peer-to-peer communications

Incorrect. Peer-to-peer communications can provide valuable data.

Learning Objective 3

Define big data in relation to social media transit



Overview

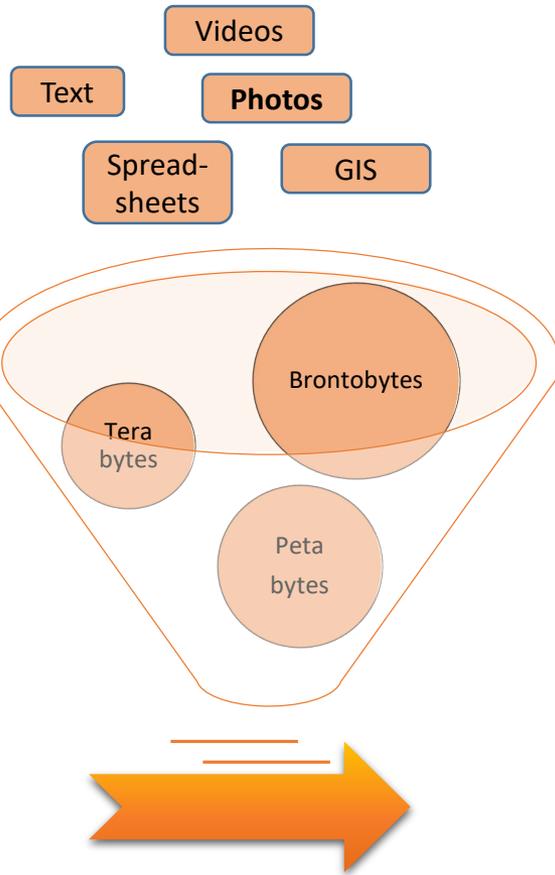
- What is big data?
- Large datasets characterized by variety, volume, and velocity
- Sources of transit-related big data include internal and external data sources
- Characteristics of social media datasets
- Social media data standards are emerging

What is Big Data?



- Large volume of data
 - Structured data
 - Unstructured data
- Difficult to process with traditional database and software techniques

Characteristics of Large Datasets



Large datasets are characterized by their variety, volume, and velocity

- **Variety**

- Multiple sources
- Multiple formats: text, photo, video, PDF, database, CSV, spreadsheets
- Structured and unstructured

- **Volume**

- Terabytes (10^{12})
- Petabytes (10^{15})
- Brontobytes (10^{27}) and upwards

- **Velocity**

- Speed required to convert inputs into outputs
- Streaming, which is continuous conversion from inputs to outputs



Examples of the 3 Vs and Transit-Related Data

Data Description	Variety	Volume – Storage	Velocity - Frequency of updates
Vehicle Location 100,000 trips per year	Structured	3.6 GB per year	50 bytes per vehicle every 5 seconds
Schedule Data (e.g., SEPTA bus)	Structured (GTFS) and compressed	21 MB	Seasonal
Video from 300 Cameras	Video	1.2 TB	Streaming
Geographic Information Files (NJT Bus)	Structured	40 MB	Seasonal

Transit-Related Big Data Includes Internal and External Sources

- Internal sources
 - Rider surveys and panels
 - Focus groups and stakeholder interviews
 - Automatic passenger counting data (APC)
 - Automated vehicle location data (AVL)
 - General Transit Feed Specification files (GTFS/GTFS-rt)
 - Electronic fare payment system datasets (EFPS)
- External data sources
 - Social media posts
 - Census files and other public datasets
 - Traffic data
 - Web pages (HTML)



Characteristics of Social Media Datasets

- Unstructured text, written in natural language
- Uncategorized
- Voluminous
- Variety of formats (e.g., JPG, GIF, MP3, MP4)

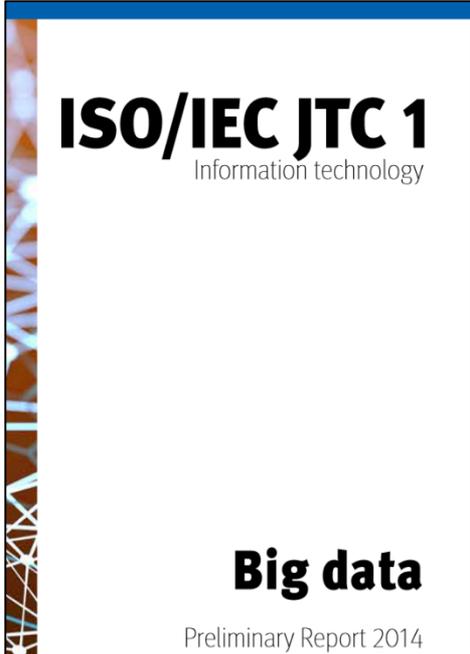


Social Media Data Standardization Challenges

Standards may be emerging, but standardization is a challenge.

- Social media is unstructured and may include natural text, images, and video.
- Social media platforms are mostly owned by private for-profit entities and data (e.g., posts) may use a proprietary format.
- Some social media have Application Programming Interfaces (APIs) for downloading data, but others have no API.

International Efforts on Big Data Standards (1 of 2)



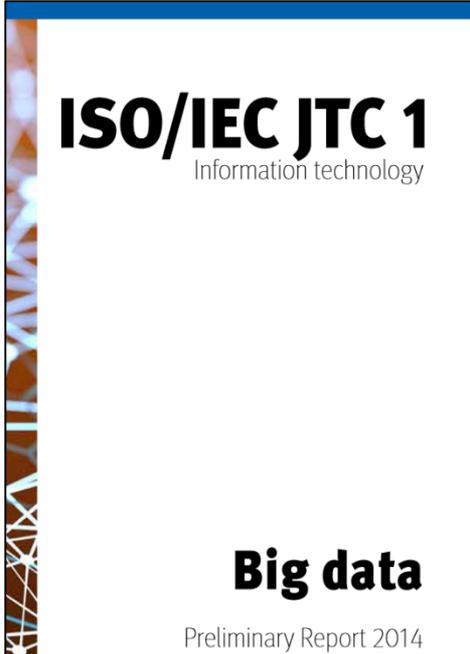
ISO/IEC JTC 1
Information technology

Big data

Preliminary Report 2014

SDO/ Consortium	Interest Area
ISO/IEC JTC 1/SC 32	Data management and interchange, including database languages, multimedia object management, metadata management and e-Business.
ISO/IEC JTC 1/SC 38	Standardization for interoperable Distributed Application Platform and Services including Web Services, Service Oriented Architecture (SOA), and Cloud Computing.
ITU-T SG13	Cloud computing for Big Data.
W3C	Web and Semantic related standards for markup, structure, query, semantics, and interchange.

International Efforts on Big Data Standards (2 of 2)



ISO/IEC JTC 1
Information technology

Big data

Preliminary Report 2014

SDO/ Consortium	Interest Area
Open Geospatial Consortium	Geospatial related standards for the specification, structure, query, and processing of location related data.
Organization for the Advancement of Structured Information Standards	Information access and exchange.
Transaction Processing Performance Council	Benchmarks for Big Data Systems.
TM Forum	Enable enterprises, service providers and suppliers to continuously transform in order to succeed in the digital economy.

ACTIVITY



Question

Which of the below is not one of the 3 V characteristics of big data?

Answer Choices

- a) Velocity
- b) Viscosity
- c) Variety
- d) Volume

Review of Answers



a) Velocity

Incorrect. Velocity refers to the speed required to convert input data into output data.



b) Viscosity

Correct! Viscosity is not one of the 3 Vs of Big Data, but a useful measure for assessing the quality of maple syrup and ketchup.



c) Variety

Incorrect. Variety refers to the diversity and inconsistency in the structured and unstructured data present in Big Data.



d) Volume

Incorrect. Volume refers to the quantity of data and growth rate.

Learning Objective 4

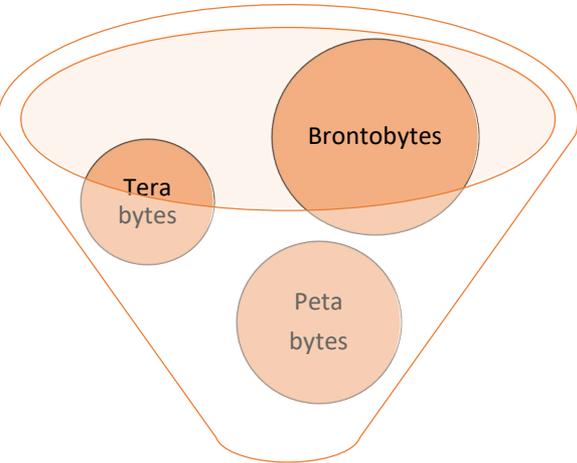
Understand the process for applying big data analytics to social media to inform transit business intelligence.



Overview

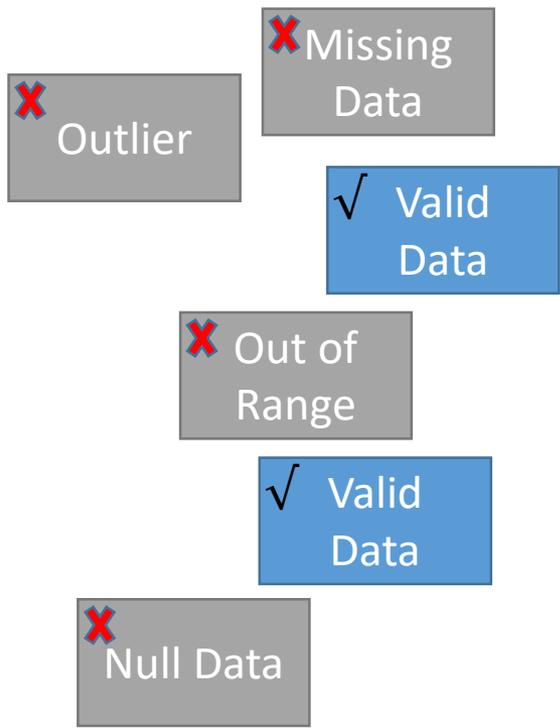
- Data acquisition
- Data preparation
- Data analysis
 - Data science techniques
- Data presentation
 - Visualization
 - Dashboards
- Other Issues
 - Policy issues
 - Technical issues

Data Acquisition

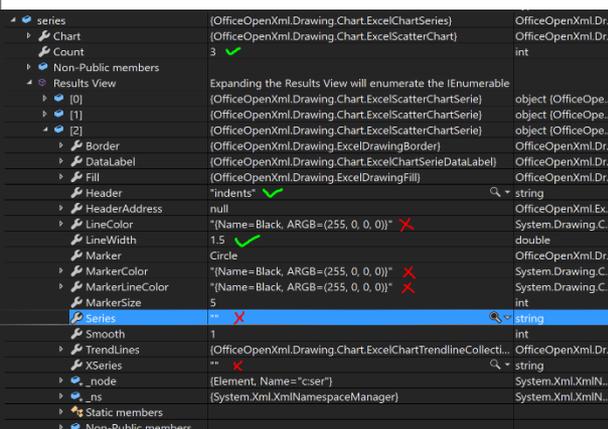


- Data acquisition is the means necessary to gather data for subsequent steps. These may include:
 - Data collection
 - Data recording of natural events
 - Data recording of human-made events
 - Data entry
- **What data do I have?**
 - Internal sources
 - External sources
- **What data do I need that I don't have?**
- Do I need:
 - To do data scraping
 - To use an Application Programming Interface
- **How much will new data cost me to acquire?**
- **What are my storage requirements**
 - Volume, security, in the cloud, in-house
- **Where do I store my data?**
 - We introduce the term “data lake”

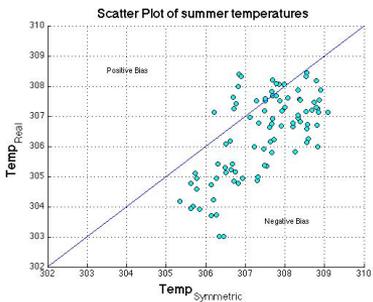
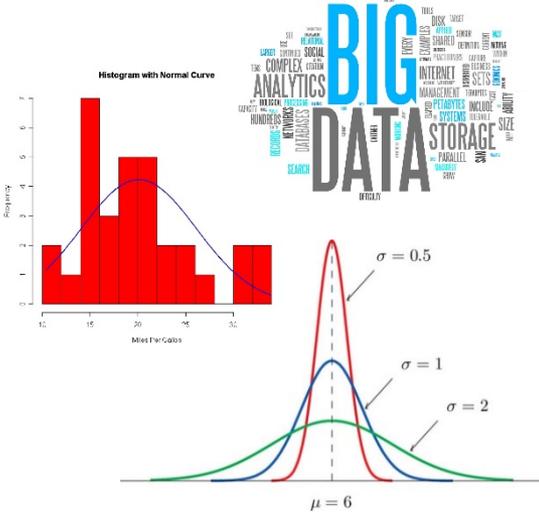
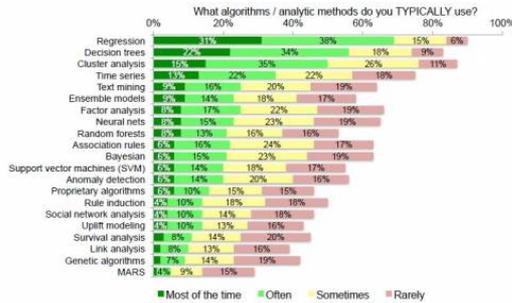
Data Preparation



- Data preparation removes data that is incomplete, incorrect, or out of range from analysis.
- Do I have the right data?
 - Granularity
 - Coverage
 - Content
 - Geographic region and data (GPS, GIS files)
 - Time frame
- If there is a standard available, this is the step to map data to the standard
- Data scrubbing and filtering occurs in this step
 - Remove outliers
 - Handle of missing data
 - Remove out of range data
 - Handle null data values
- Define any rules for sentiment analysis, topic maps, and linkages between disparate data sets.



Data Analysis



- Data analysis is the interpretation of relationships between data to gain insights about a problem or solution.

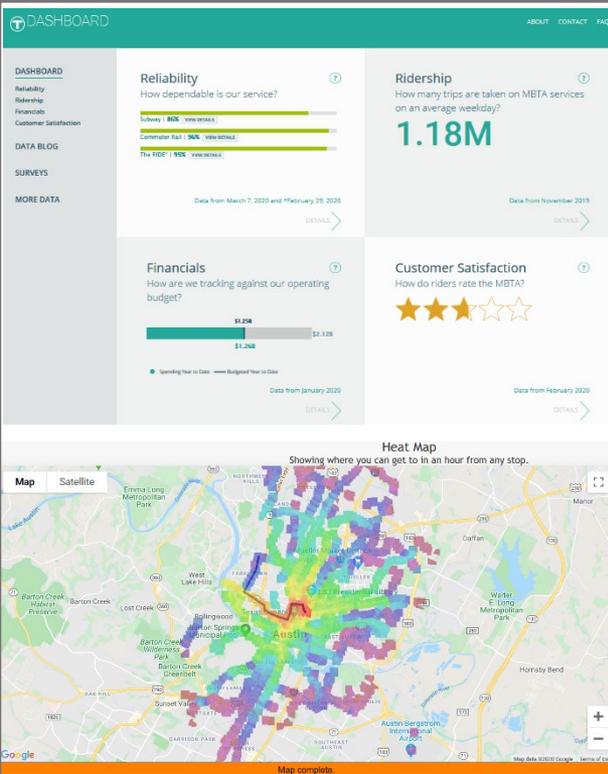
- Data analysis techniques include

- Data mining
- Data visualization
- Topic maps
- Sentiment analysis
- Data similarity analysis
- Stochastic analysis
- Data correlation

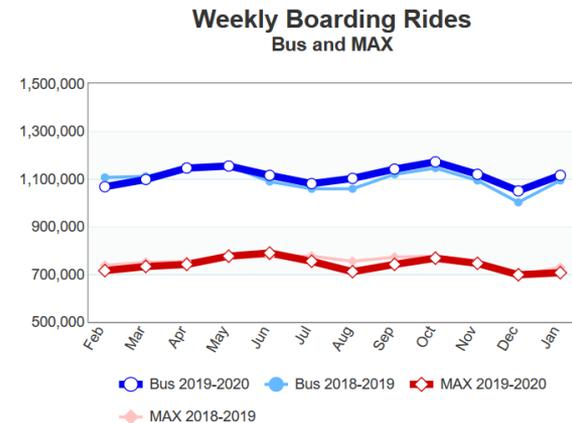
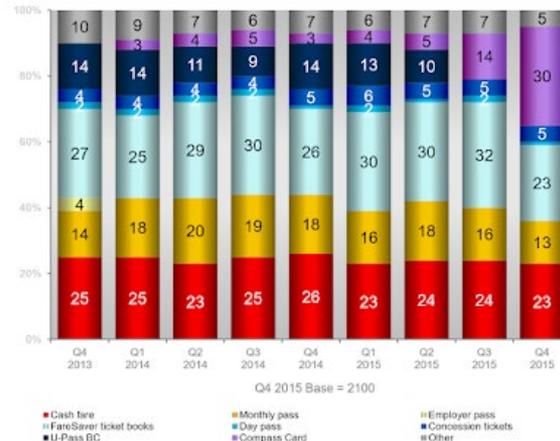
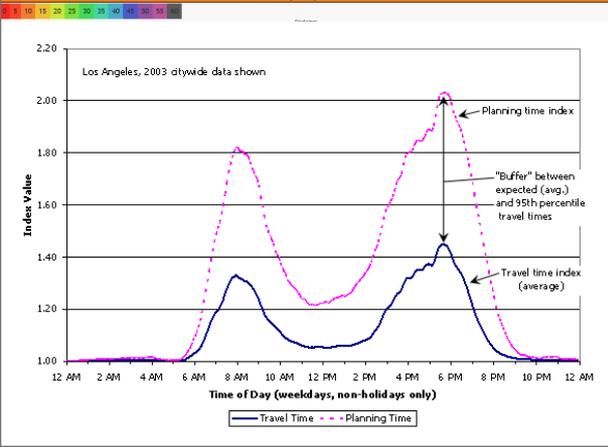
- Artificial intelligence and machine learning

- Image processing
- Facial recognition
- Automated license plate recognition (ALPR)
- Predictive analytics

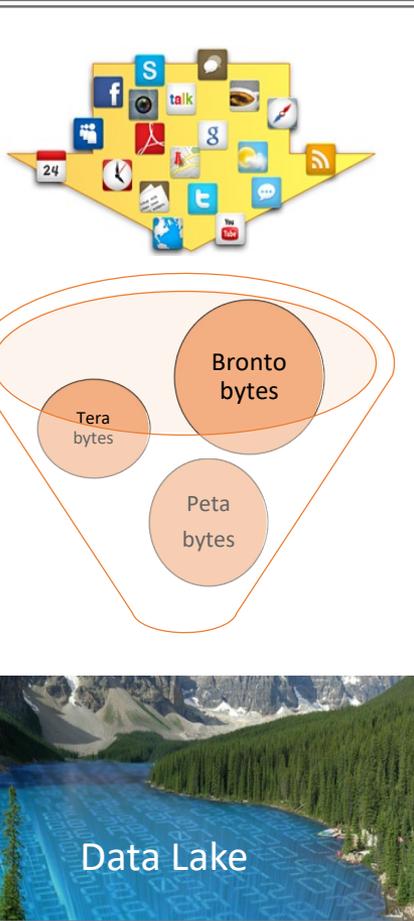
Data Presentation



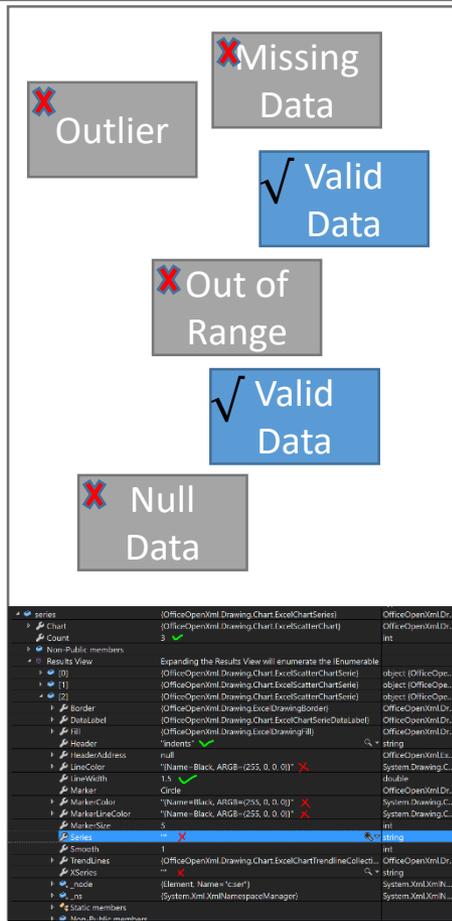
- Data presentation is the process of using the results of analysis to provide an explanation or make a claim about the data.
- Agency dashboards draw data from multiple sources to share key performance indicators:
 - Ridership
 - Service performance
 - Financial
 - Customer satisfaction
 - Maintenance records
 - Electronic fare payment



Big Data Process Steps Summary

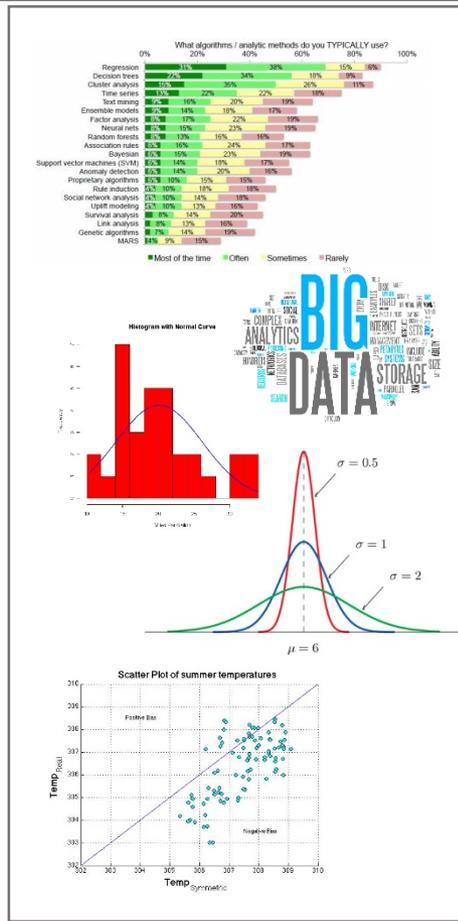


A funnel-shaped graphic representing data acquisition, filled with various social media and technology icons. Below it, a landscape image of a lake with the text "Data Lake" overlaid.

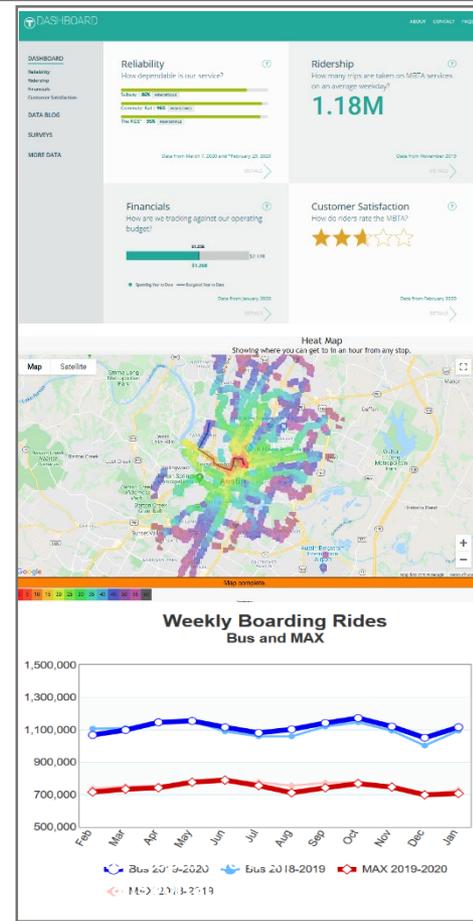


A diagram showing data preparation steps. It lists "Missing Data", "Outlier", "Valid Data", "Out of Range", "Valid Data", and "Null Data" with corresponding checkmarks and X marks. Below is a screenshot of a data table with various columns and values, some marked with red X's.

series	(OfficeOpenXml.Drawing.Chart.ExcelChartSeries)	OfficeOpenXmlDr...
Chart	(OfficeOpenXml.Drawing.Chart.ExcelChartSeries)	OfficeOpenXmlDr...
Count	3	int
Non-Public members		
Results View	Expanding the Results View will enumerate the IEnumerable<object> (OfficeOpt...	object (OfficeOpt...
0	(OfficeOpenXml.Drawing.Chart.ExcelChartSeries)	object (OfficeOpt...
1	(OfficeOpenXml.Drawing.Chart.ExcelChartSeries)	object (OfficeOpt...
2	(OfficeOpenXml.Drawing.Chart.ExcelChartSeries)	object (OfficeOpt...
Border	(OfficeOpenXml.Drawing.ExcelDrawing.Border)	OfficeOpenXmlDr...
DataLabel	(OfficeOpenXml.Drawing.Chart.ExcelChartSeries.DataLabel)	OfficeOpenXmlDr...
Fill	(OfficeOpenXml.Drawing.ExcelDrawing.Fill)	OfficeOpenXmlDr...
Header	Header	string
HeaderAddress	null	OfficeOpenXmlDr...
LineColor	"Name-Black, ARGB=(255, 0, 0, 0)"	System.Drawing.C...
LineWidth	1.5	double
Marker	Circle	OfficeOpenXmlDr...
MarkerColor	"Name-Black, ARGB=(255, 0, 0, 0)"	System.Drawing.C...
MarkerLineColor	"Name-Black, ARGB=(255, 0, 0, 0)"	System.Drawing.C...
MarkerSize	5	int
Series	(OfficeOpenXml.Drawing.Chart.ExcelChartSeries)	OfficeOpenXmlDr...
Smooth	1	int
TrendLines	(OfficeOpenXml.Drawing.Chart.ExcelChartTrendlineCollect...	OfficeOpenXmlDr...
XSlices	1	int
YAxis	(System.Xml.XmlElement) Name="YAxis"	System.Xml.XmlEl...
YAxisTitle	(System.Xml.XmlElement) Name="YAxisTitle"	System.Xml.XmlEl...
Static members		



A collection of data analysis charts. Top: A horizontal bar chart titled "What algorithms / analytic methods do you TYPICALLY use?" showing usage percentages for various methods like Regression, Decision trees, etc. Middle: A word cloud with "BIG DATA" as the largest word. Bottom: A histogram with normal curves for different standard deviations (σ = 0.5, 1, 2) and a scatter plot of summer temperatures for Foster and Tropic Basins.



A dashboard titled "DASH-BOARD" showing various metrics. Top: "Reliability" and "Ridership" sections with progress bars and a large "1.18M" figure. Middle: "Financials" and "Customer Satisfaction" sections with bar charts and star ratings. Bottom: A "Heat Map" showing data distribution on a map and a "Weekly Boarding Rides Bus and MAX" line chart comparing data from 2019-2020.

Data Acquisition

Data Preparation

Data Analysis

Data Presentation

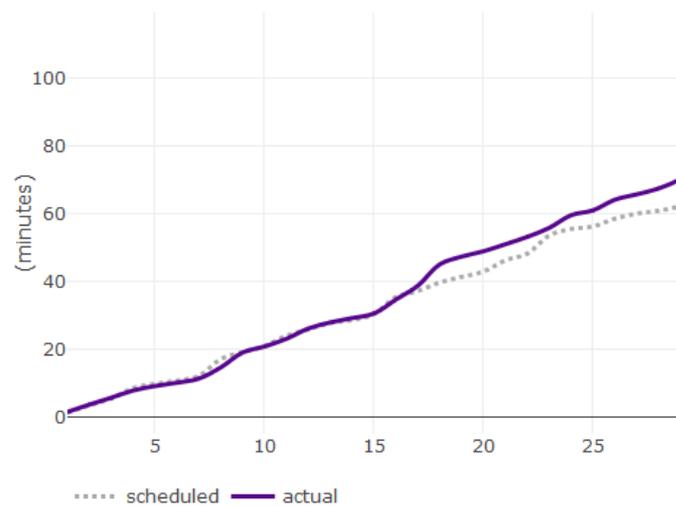
Data Presentation Example: MBTA Dashboard



- MBTA has an online dashboard for key performance indicators
- Supports transparency in reporting for internal and external audiences
- Supports drill down by mode, line, and route to get a snapshot of service performance.
- This dashboard does not include social media posts.
- The URL is in the Student Supplement.

Data Presentation Example: busstat.nyc

CUMULATIVE TRAVEL TIME ACROSS STOPS



EXCESS WAIT
TIME

2.0 mins

ROUTE
LATENESS
FACTOR

39.4 %

AVERAGE
SPEED

8.0 mph

- busstat.nyc measures and displays performance for New York City buses.
- Project is in beta as of January 2020.
- Proposed metrics join data from multiple sources to generate performance indicators that reflect customer experience and agency progress toward meeting goals.
- Route lateness factor compares actual trip time to scheduled trip time. No social media posts were included.
- Project developed by the NYU Center for Urban Science and a capstone project of the master program sponsored by TransitCenter
- The URL is in the Student Supplement.



Policy Issues

- Protecting user privacy
- Data security
- Regulatory environment and limitations/policy of government agencies use of social media
- Understanding how well social media data represents agency customer base
- Analyzing social communications in multiple languages

Technical Issues

- A data lake may be partitioned into “data ponds” to:
 - Limit access
 - Share data resources with another agency
 - Provide a means of data sharing between agencies.
 - A regional lake may provide ponds for separate transit properties
- Open source/open data tools
 - Need to consider whether adequate technical support and security are available
- Resource requirements (e.g., skills, storage, hardware, licensing, in-house vs. contracted)

ACTIVITY



Question

Which of the below is not a step described in Big Data processing?

Answer Choices

- a) Data Preparation
- b) Data Field Quantization
- c) Data Analysis
- d) Data Acquisition

Review of Answers



a) Data Preparation

Incorrect. Data preparation is the step of removing data that is incomplete, incorrect, and/or out of range from analysis.



b) Data Field Quantization

Correct! Data field quantization evaluates elements of the General Relativity Theory to prove gravity exists and is the basis for the general rule that buses will roll instead of fly.



c) Data Analysis

Incorrect. Data analysis is the interpretation of relationships between data to gain insights about a problem or solutions.



d) Data Presentation

Incorrect. Data presentation is the process of using the results of analysis to make a case or explanation about data.



Learning Objective 5

Incorporate findings to support business intelligence with data-driven decisions

Working with Social Media

- Social media posts from transit customers, stakeholders, and others (inbound communications) can provide unfiltered feedback.
- Social media posts use natural language, which requires special analytical techniques to create meaningful datasets.
- Posts usually include usernames, which must be removed during analysis to protect privacy.
- Some transit agencies restrict use of social media by staff.
- Social media users may not be representative of all transit customers.



EXAMPLE

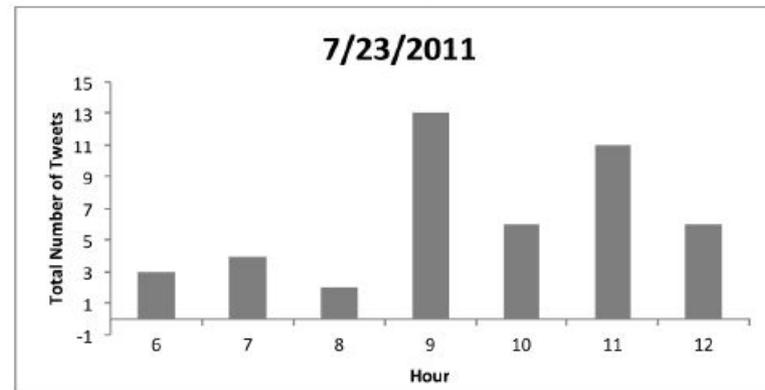
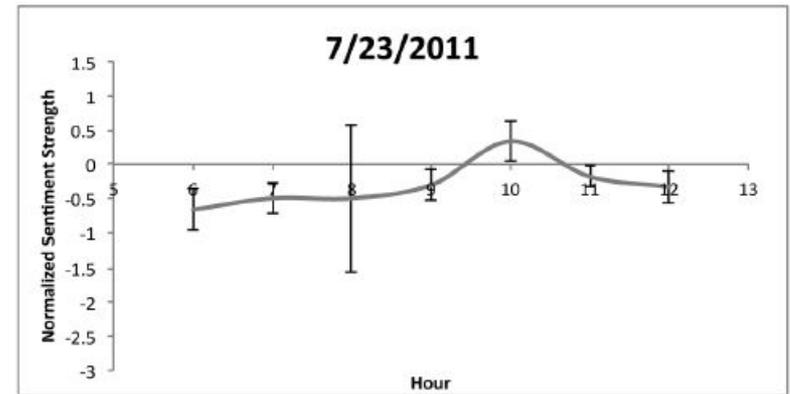
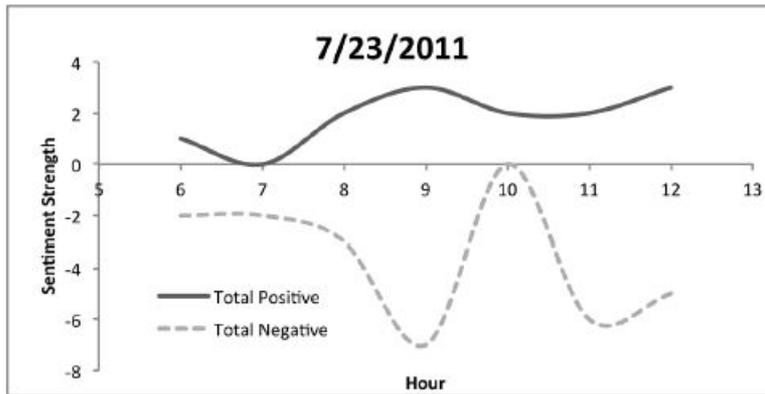
- Chicago Transit Authority (IL)
- San Diego Metropolitan Transit System (CA)
- Transport for London (UK)
- Metro Transit (MN)



Measuring Customer Sentiment

- In one of the first papers on the topic, researchers analyzed tweets that mentioned the Chicago Transit Authority to better understand customer sentiment.
- Researchers assembled a dataset of Twitter posts that mentioned CTA or individual lines.
- Analysis determined that customers were more likely to express negative sentiments toward a situation than positive sentiments.

Negative tweets spiked at 9 AM on July 23, 2011.



Combat Fare Evasion

- San Diego Metropolitan Transit System used big data to help combat fare evasion on trolleys.
- Trolleys use barrier-free honor system to collect fares. Customers tap smartcards to fare validators on the platform.
- MTS contracted with a consultant to analyze fare payment patterns.



Combat Fare Evasion

- Analysis incorporated multiple data sources.
 - GTFS showed vehicle location.
 - Fare validators showed smartcard taps before boarding.
 - Automatic passenger counters calculated boardings per station.
- Data analysis correlated farecard taps with passenger counts and vehicle arrivals to determine locations for additional fare enforcement.
- Social media was not a data source for this analysis.



Optimizing Advertising

- Researchers tested a methodology for analyzing geotagged social media posts in Transport for London Underground stations to optimize advertising campaigns.
- Tweets were analyzed and categorized based on topics of interest (e.g., sports, entertainment).
- Information was intended to provide guidance for advertising campaigns at different stations.



Locating Bus Shelters

- Metro Transit in Minneapolis/St. Paul uses big data analytics.
- Strategic Initiatives Department draws on data from multiple sources to support data-driven decision making.
 - How to allocate resources for bus shelters and amenities?
 - How to improve on-time performance?
 - How to design a transit network to best meet customer needs?

Locating Bus Shelters

- Data sources
 - Customer survey
 - Facilities
 - Ridership
 - Demographics
- Equity-focused measures were developed to inform decisions.
- Data sources do not include social media.



ACTIVITY



Question

Based on these examples, analyzing social media data helped inform agency decisions about which of the following?

Answer Choices

- a) Where to upgrade bus shelters
- b) How to understand customer sentiment
- c) Where to add fare enforcement
- d) How to report non-fare revenues

Review of Answers



a) Where to upgrade bus shelters

Incorrect. Agency did not consider social media posts.



b) How to understand customer sentiment

Correct! Researchers analyzed social media posts to assess CTA customer sentiment.



c) Where to add fare enforcement

Incorrect. Agency did not use social media to solve problem.



d) How to report non-fare revenues

Incorrect. None of the examples focused on non-fare revenues. Social media is not a source of this data.

Module Summary

Learned how transit operators can use business intelligence tools to make data-driven decisions

Saw examples of agency-generated and customer-generated social media posts

Learned about potential sources of big data for use in transportation analysis

Reviewed process for applying big data analytics to social media to inform transit business intelligence

Reviewed examples of using big data to support business intelligence

Thank you for completing this module.

Feedback

Please use the Feedback link below to provide us with your thoughts and comments about the value of the training.

Thank you!