Data Matters:
Making The Most of Data

WESTERN INTERGOVERNMENTAL AUDIT FORUM MEETING

September 5, 2018

Nelson, Kat, and Ron

Nelson

City & County of San Francisco
Supervising Auditor in
IT & Systems Audit team

Recent data worked with:

- Purchase card transactions for continuous auditing
- Procure to pay data for continuous auditing
- Continuity of operations plans

Kat

City & County of San Francisco
Supervising Auditor in
Construction & Performance Audit team

Recent data worked with:

- Below market-rate rental unit pricing, marketing, and tenant
- City employee payroll
- Construction project budget, expenditures, and schedule
- IT helpdesk
- Qualitative and quantitative survey data
- Best practice and academic research articles
- Program application

Ron

Arizona Auditor General
Data Analysis and Strategic Applications Team Manager

Recent data worked with:

- State agencies
- School districts
- Universities
- Counties

Outline

- What is Data Analytics?
  - Data analytics defined
  - General methodology
  - Project-based analytics versus continuous auditing
- Getting Data
  - Identifying needed data and available data
  - Data extraction and data types
  - Quickly assessing the quality of available data
  - Bridging the gap between your data need and what is available
- Data Analytics Tools

What is Data Analytics?

PROCESS

of *inspecting*, *cleaning*, *transforming*, & *modeling* data to gain **insight**, discover *meaningful relationships*, inform conclusions, and support decision-making.

Analytics can provide clues to help you go from **condition** to **root cause**

Diving in to Data Analytics

Prepare	*Benefit*	*Invest*
Support and buy-in from leadership	Perform, evaluate results, resolve	Methodology must be effective & sustainable
Collaboration with the owners of data	Increase coverage, quality, and impact	Acquiring tools
Start small with a pilot project	Strengthen skills of auditors	Learning curve

Typical Data Analytics Methodology

Plan

- Clear Objectives; Obtain Buy-In; Reasonable Targets

Acquire & Understand

- Initial Data Samples; Obtain Owner Insights; Verify Data Quality; Choose The Best Analysis Tool

Analyze

- Develop Hypothesis; Prepare Summary; Adjust Scope As Needed

Validate

- Translate & Discuss Results; Adjust Testing As Needed; Obtain Clear Feedback

Report

- Formally Present Findings; Provide Complete Transparency; Agree On Recurring Impact; Define Action Plan; Implement Sustainable Solution

Continuous Controls Auditing

Continuous auditing ranges from continuous control assessment to continuous risk assessment – all activities on the control-risk continuum.

--The IIA, GTAG 3: Continuous Auditing: Implications for Assurance, Monitoring, and Risk Assessment

Purpose	Manage activities ranging from continuous control assessment to continuous risk assessment.
Automated Tools	Using automated tools is one of the methods of continuous auditing.
Frequency	Auditing frequency includes Daily, Weekly, Monthly, etc.
Repeatable Process	Be consistent with a repeatable process to ensure quality and accuracy of the data.

Continuous Monitoring – Purchasing Cards

Invalid cardholder

(no matching employee or terminated employee)

Declined and disputed transactions

Ghost card activities

Duplicate cardholders

(by employee ID or address)

Potential duplicate reimbursements

(gas with mileage or P-Card with an AP purchase)

Even/small dollar amount transactions

Spending limits on transactions

(lavish hotel stays, dinners, etc.)

Duplicate purchases

(same merchant same amount)

Weekend and holiday transactions

Split purchases

Suspicious keyword in the transaction description

New cardholder watch list

Suspicious merchant classification codes

More Continuous Audit Examples

General Ledger

- Unauthorized journal entry (JE)
- JEs by unauthorized users
- Duplicate JEs (same account/amount, same JE number/amount)
- Split JEs (single JE/multiple accounts, multiple JEs/single account)
- Segregation of duties (park vs. post, post vs. create account)
- Dormant accounts
- Even dollar JEs
- Suspicious keyword in JE description
- Duplicate GL accounts based on the account description

Payroll

- Duplicate employees (same bank account or address)
- Pay ceases when employee separated
- Exempt hours worked vs. standard hours
- Non-exempt hours worked vs. expected hours
- Hours worked vs. hours paid
- Employee start date after paycheck date
- Terminations within 14 days of hire
- Invalid pay rates (actual/calculated vs. master file)
- Excessive gross pay
- 401k annual contribution limit, catch-up contribution limit and catch-up
- Age limit
- Job record deletions (data corrections not using effective date)
- Overtime limit exceeded
- Leave accrual caps exceeded

Purchase to Pay

- Split requisitions and POs
- Stale requisitions and POs
- Segregation of duties
- PO date after invoice date
- Invoice number sequence not continuous
- Goods received quantity vs. invoice quantity
- Fraud of employee posing as vendor by match name and address
- Duplicate vendors (by name, address, bank account number)
- Duplicate purchases (same vendor with either same invoice number or same amount and near date, same GL account)

Getting Data

Identifying *NEEDED* Data

❖ Incorporate consideration of data needed from the beginning of audit planning

#	Audit Objective	Researchable Question(s)	What should be and who says so? (Criteria & Source)	What could go wrong? (Risk or Negative Effect)	Information Required to Answer Question & Potential Sources
EX.	Assess the strategic planning process of the SFMTA IT function.	Does the IT strategic plan align with SFMTA departmental strategic plan?	COBIT	Waste of resources in the IT unit on projects unrelated to SFMTA mission; greater likelihood of waste and abuse	IT Strategic plan from IT Management; SFMTA strategic plan from department website
EX.	Assess adequacy of staffing levels of the SFMTA IT division.	Does the Service Management unit have enough staff to respond to help desk tickets within the time frames required by the SLA?	Service level agreement (SLA) between IT division and SFMTA operational units	Urgent IT problems not being addressed promptly, vulnerability to cyberattacks when security related requests not addressed quickly, loss of productivity	Helpdesk system data for 5 years for trend analysis; SLA

Identifying *AVAILABLE* Data

- Process walkthroughs
  - Look for data outside of systems
    - Excel spreadsheets and Access databases
    - Physical files
- System walkthroughs
  - Identify all systems with relevant data
  - Identify the source of all data within that system
    - Is some data fed from another system?
    - What is manually inputted?
    - What is the source of manually inputted information?

Data Types & Structures

The majority of data is organized into tables.

A table can be thought of as a two dimensional matrix of data.

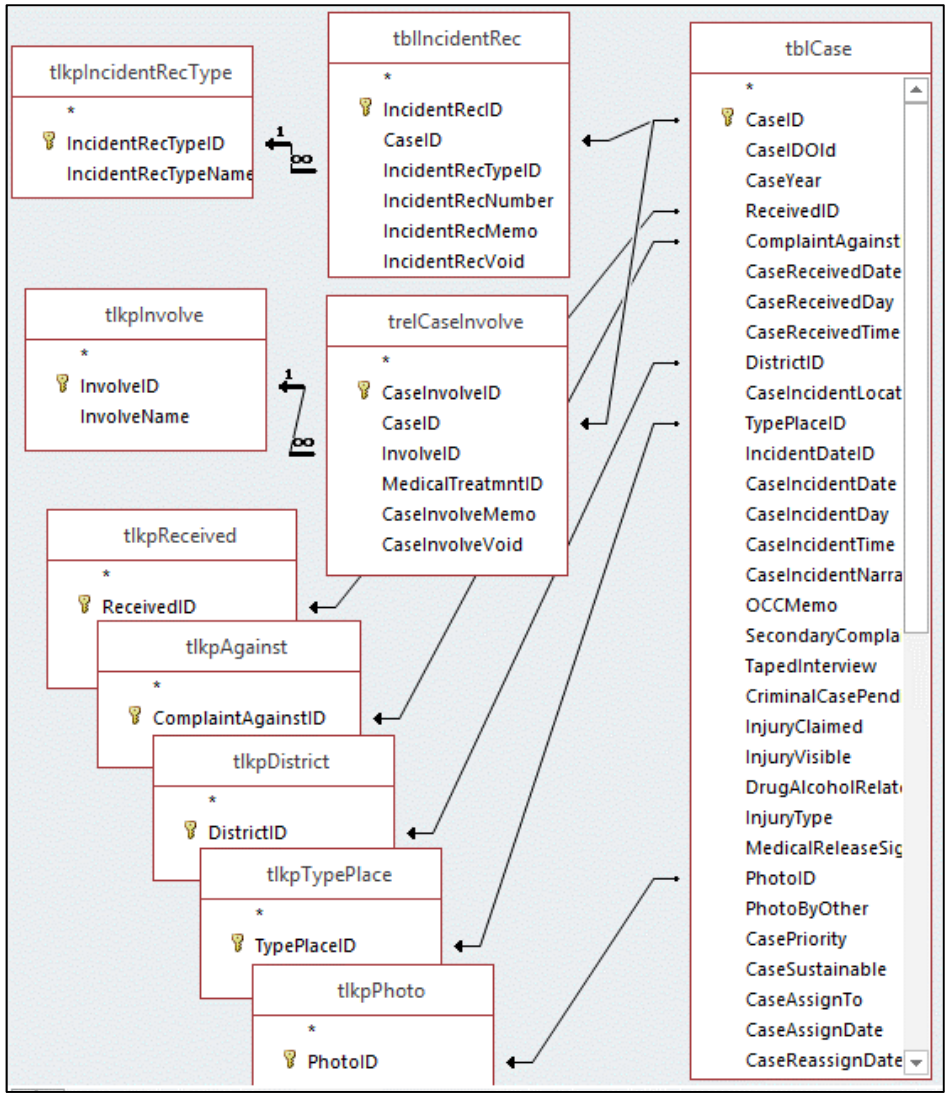
- Each row represents a single record.
- Each column represents a data field.

Each data column, or field, may have a different data type. Furthermore, data types determine how data is interpreted, and also what data format is considered valid.

- For example, data can be a date, a number, or plain text.
- Invalid data in a table is often a sign of some other problem.

Each record in a table can be related to data in another table through a unique identifier, like an employee, customer, or transaction ID.

- When one table uses this identifier to reference records in another table, this is called a relational database.
- Relational databases are a very useful way to organize data.
- Many databases are built using some kind of relational database format.



Data Extraction

- **Extract strategy depends on the source types, data types and structures**
- **Internal vs. External**
- **Production vs. Staging (or Data Warehouse)**
- **Flat Files vs. Relational Database Management Systems (RDBMS)**
- **Data Stewards vs. Database Administrators (DBA)**
- **Simple rules for moving source data to the target:**
  - Ad hoc queries of the source system(s)
  - Analysis of a small sample set of source data
  - Obtain knowledge or documentation of how the source system(s) work
  - Obtain knowledge of how the target system works
  - Sufficient access to source systems
  - Requires lots of technical input and little business input

Data Extraction Readiness Check

Regardless of structure, type, or format, source data intended for extraction should be validated in terms of the following key attributes:

- Relevance: Is it relevant for its intended purpose?
- Accuracy: Is it correct and objective, and can it be validated?
- Integrity: Does it have a coherent, logical structure?
- Consistency: Is it consistent and easily to understand?
- Completeness: Does it provide all the information required?
- Validity: Is it within acceptable parameters for the organization?
- Timeliness: Is it up to date and available whenever required?
- Accessibility: Can it be easily accessed and exported to the target application?

Quickly Assess the Quality of *AVAILABLE* Data

Jump to...

System Understanding

1. **System Name** -
2. **System Purpose** –
3. **Implementation** - Was the system created or purchased?

   ☐ Created by in

   ☐ Purchased from in
4. **Administration** - Who is responsible for system administration and maintenance?

Data

When relevant, obtain ***screen shots*** of the system to support

Information Request

The following inf
should be reques
the system walkth
reference links). A
to this walkthrou
after reviewing th
documents:

- ☑ System user ma
- ☑ System design s
- ☑ Data definitions
- ☑ List of system ro definitions inclu areas of the syst access to and w have read/write/ access
- ☑ Complete list of including generi administrative a the accounts rol system
- ☑ A list of users sp

8. What controls are built in the ***system*** to ensure the data is entered accurately?

   ☐ Yes ☐ No - Records have a unique identifier (system does not allow duplicates)

   Data label for unique identifier:

   ☐ Yes ☐ Partially ☐ No - Key fields are mandatory:

   ☐ Yes ☐ Partially ☐ No - All fields only allow valid entries (text in text fields, numbers in number fields, etc.)

   ☐ Yes ☐ Partially ☐ No - Fields with restricted options use drop-down menus

   ☐ Yes ☐ Partially ☐ No - Logic checks that disallow or raise flags for unexpected values (8 digits allowed in a phone number field; 25 allowed as a supervisor district number; 1,000,000 allowed when the average value is 500, etc.)

   ☐ Yes ☐ Partially ☐ No - System flags potential duplicate entries (two purchases from the same vendor at the same cost within 1 day of each other; two applicants with the same address, etc.)

   ☐ Yes ☐ Partially ☐ No – Other:

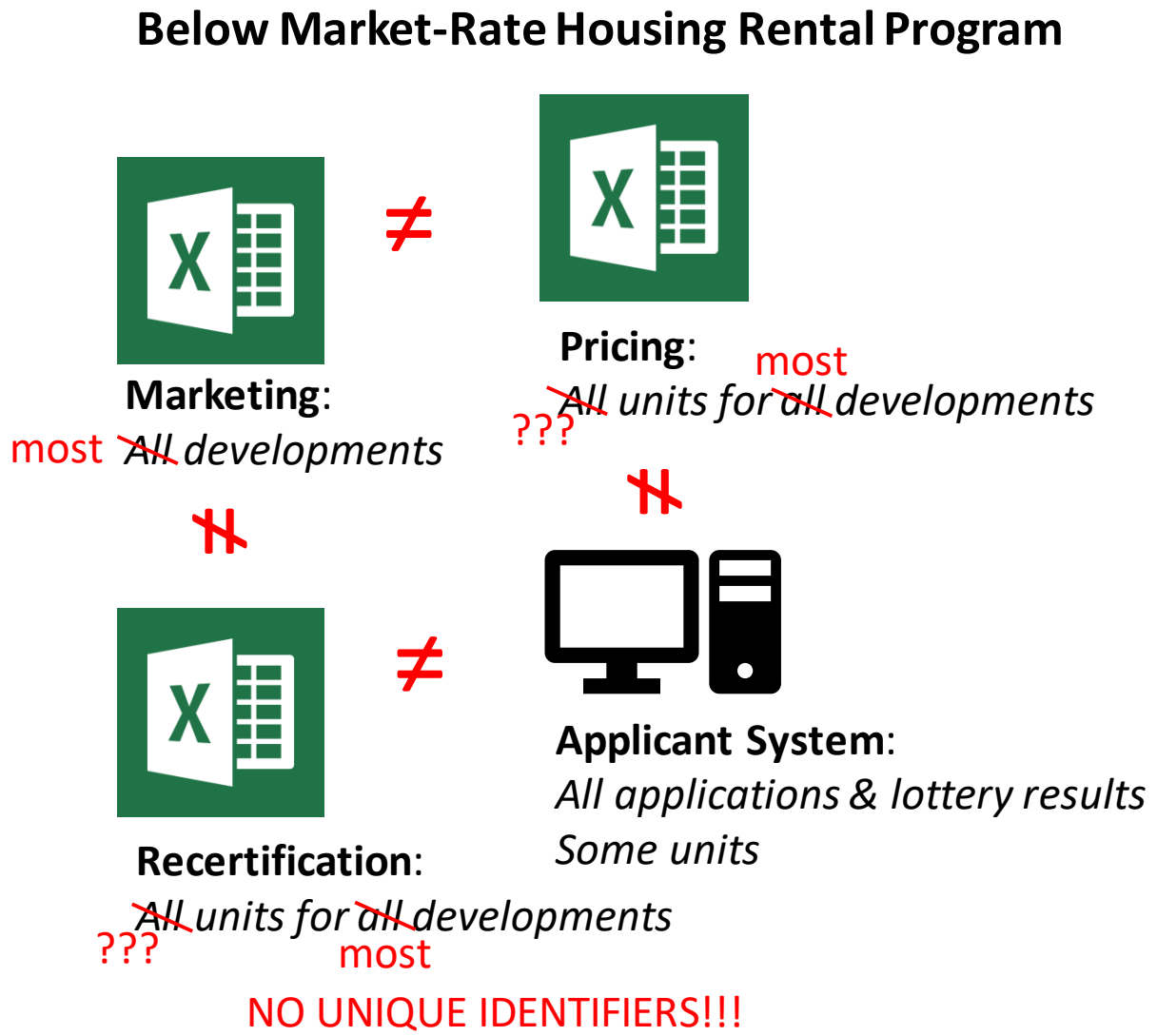
9. What ***human controls*** are built in to the process to ensure that data entered into the system is accurate?

   ☐ Yes ☐ Partially ☐ No – Entries flagged by the system are reviewed by department staff

Bridging the gap between NEED and AVAILABLE

Identify the source of the gap:

- ❖ No data
- ❖ Inaccurate data
- ❖ Incomplete data
- ❖ Decentralized data
- ❖ Data in a format that inhibits analytics



Bridging the gap between NEED and AVAILABLE

Identify the source of the gap:

- ❖ No data
- ❖ Inaccurate data
- ❖ Incomplete data
- ❖ Decentralized data
- ❖ Data in a format that inhibits analytics

Determine what it would take to get around the gap:

- ❖ Is there another source for the data or combination of sources?
- ❖ Can the existing data be used with more limited scope?
- ❖ Can incomplete data be made complete?
- ❖ Can inaccurate data be corrected?
- ❖ Can the data be created?

Assess the value of the effort of getting around the gap.

Below Market-Rate Housing Rental Program

Housing Office

Applicant System

Marketing

Pricing

Recertification

Planning Department

Proposed Developments:
All developments

Building Inspection

Buildings Completed:
All developments, all units

Data Analytics Tools

Picking the Right Tools

Design/ Implementation

- Single user vs. multiple users
- Desktop vs. server
- Data vs. visualizations
- Securing data

Expertise/ Training

- In-House expertise
- Online training
- Consulting

Cost

- License fees / Maintenance
- Specialized hardware

Analytic Arsenal

- **Spreadsheets** (Microsoft Excel)
- **Desktop Databases** (Microsoft Access)
- **Audit Software** (ACL)
- **Server Databases**
  (SQL Server, including Reporting Services)
- **Visualizations** (Microsoft Power BI)
- **Qualitative Analysis** (Atlas.ti)

Spreadsheets

(Microsoft Excel)

Pros

- Easy to use
- Create simple lists
- Conditional Formatting
- Run calculations and statistical comparisons
- Pivot tables
- Charts
- Macro scripting (VBA)

Cons

- Limited to 1,048,576 rows and 16,000 columns
- Little to no audit trail
- Problematic as data grows
- Security
- Viruses attached through macros
- Multiple copies

Desktop Databases

(Microsoft Access)

Pros

- Readily available
- Integration
- Relational database
- Multi-user
- Larger storage capacity
- Data can be updated
- Queries
- Simple Report creation

Cons

- Limited to individuals or small groups
- Little or no audit trail
- Maintenance and corruption
- Limited Reporting/Visualization
- Multiple copies
- Database 101
- Security

Audit Software
(ACL)

- Import wide variety of data types
- Read-only data
- Robust audit/analysis functions
- Performance
- Automation
- Logging

Pros

- Lots of disk space required
- Data cannot be directly cleansed
- Cumbersome relationships and indexes
- Reporting good but limited
- Cost
- Security

Cons

Database Servers
(SQL Server)

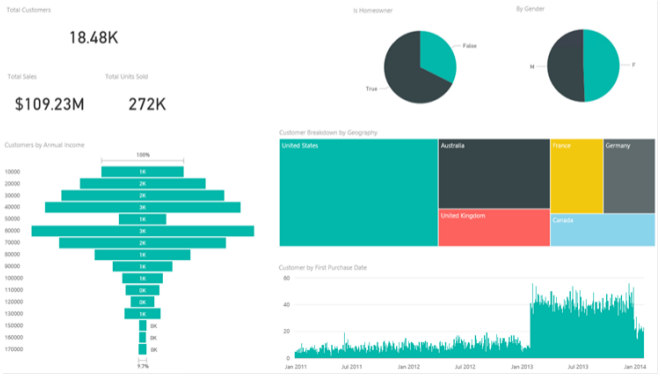
Pros

- Capacity
- Performance
- Integration
- Dynamic data
- Views (joined datasets)
- Multi-user
- Granular security
- Transformations
- Automation

Cons

- Requires IT administration and maintenance
- Learning curve (database management, admin, design)
- GUI limited (coding needed)
- Cost
- Runs on Windows-based servers

Database Servers
(Reporting Services)



- Interactive reports allow users to select their criteria from drop-down lists
- Easy export for further analysis with multiple file formats
- Web based
- Subscriptions
- Role-based security model

Pros

- Limited data sources (SQL Server only)
- Learning curve involved

Cons

Reporting Services

Leadership Competencies
LEAP Management Report
Officewide (2016)

Officewide

		Competency Ratings	Responses
LEADING CHANGE	1. Creativity & Innovation	6.44	103
	2. External Awareness	6.11	103
	3. Flexibility	7.47	103
	4. Resilience	7.33	103
	5. Strategic Thinking	5.37	103
	6. Vision	5.2	103
LEADING PEOPLE	7. Conflict Management	6.55	103
	8. Leveraging Diversity	6.73	103
	9. Developing Others	6.87	103

Reporting Services

Total number of initial deficiencies:	20
Number of significant deficiencies:	13
Compliance Status	Compliant

Our opinion on the District's compliance status is based on our review of the deficiencies cited by the District's auditors in the USFR Compliance Questionnaire, auditor's reports on internal control and compliance, auditor's opinion on the financial statements, management letter, and financial statements, (as applicable) and the District's compliance history.

Any additional considerations:

	Progress in correcting previously cited deficiencies						
Area (highlight area in yellow if emphasizing in MC letter)	**# Prior Year Deficiences**	**# Corrected**	**# Not Fully Corrected**	**# New**	**# Current Year Deficiencies**	**In the current year, the District has:**	**Increase Risk of Fraud** (only on Noncompliance or Marginal that are "Nearly Noncompliance")
ACCOUNTING RECORDS	2	1	1	1	2	Not Improved	
AUXILIARY OPERATIONS AND EXTRACURRICULAR ACTIVITIES FEES TAX CREDIT FUNDS							
BUDGETING				1	1	Worsened	
CASH AND REVENUES	5	3	2	1	3	Not Improved	
CLASSROOM SITE FUND	1	1				Improved	
COOPERATIVE AGREEMENTS AND REGIONAL SERVICES							
CREDIT CARDS AND P-CARDS							
EXPENDITURES	1	0	1		1	Not Improved	
FINANCIAL REPORTING							
FOOD SERVICE FUND							
GENERAL LONG-TERM DEBT							

Visualization Tools
(Power BI)

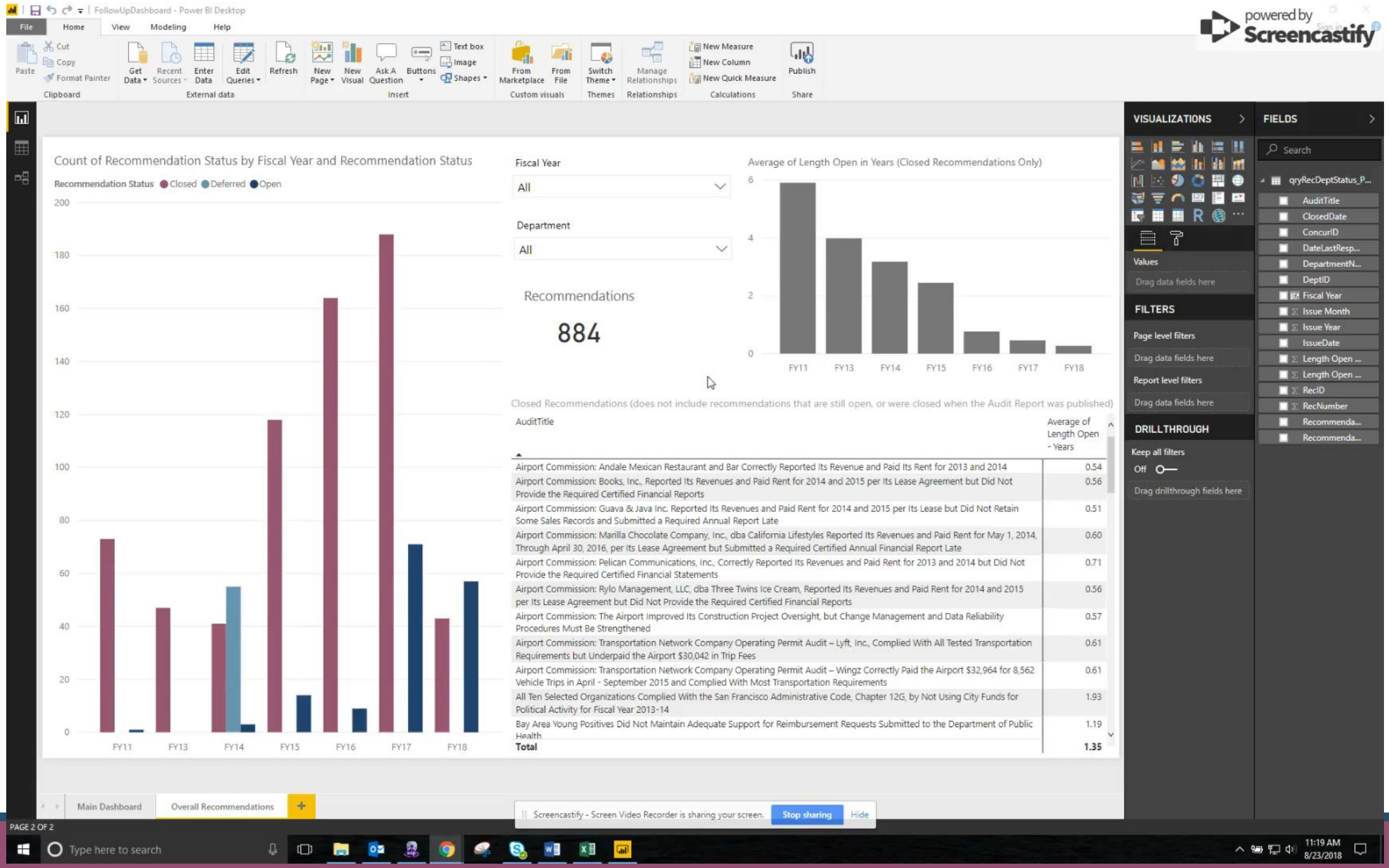
Pros

- Easy to get started and use
- Imports from many sources
  - Live links(fast/dynamic)
  - Imported(robust functionality)
- Interactive visualizations
- Easy to share
- Resources and community

Cons

- Printing
- Formatting limited
- If data changes, recreate.
- Table relationships limited
- Only sharable in same email domain
- Cannot mix imported and live links data
- Live links limited functionality
- Sluggishness with large data sets (1GB data import limit)

Power BI


FollowUpDashboard - Power BI Desktop
File
Home
View
Modeling
Help
Count of Recommendation Status by Fiscal Year and Recommendation Status
Recommendation Status Closed Deferred Open
200
180
160
140
120
100
80
60
40
20
0
FY11
FY13
FY14
FY15
FY16
FY17
FY18
Fiscal Year
All
Department
All
Recommendations
884
Average of Length Open in Years (Closed Recommendations Only)
6
4
2
0
FY11
FY13
FY14
FY15
FY16
FY17
FY18
Closed Recommendations (does not include recommendations that are still open, or were closed when the Audit Report was published)
AuditTitle
Average of Length Open - Years
Airport Commission: Andale Mexican Restaurant and Bar Correctly Reported Its Revenue and Paid Its Rent for 2013 and 2014 0.54
Airport Commission: Books, Inc., Reported Its Revenues and Paid Rent for 2014 and 2015 per Its Lease Agreement but Did Not Provide the Required Certified Financial Reports 0.56
Airport Commission: Guava & Java Inc. Reported Its Revenues and Paid Rent for 2014 and 2015 per Its Lease but Did Not Retain Some Sales Records and Submitted a Required Annual Report Late 0.51
Airport Commission: Marilla Chocolate Company, Inc., dba California Lifestyles Reported Its Revenues and Paid Rent for May 1, 2014, Through April 30, 2016, per Its Lease Agreement but Submitted a Required Certified Annual Financial Report Late 0.60
Airport Commission: Pelican Communications, Inc., Correctly Reported Its Revenues and Paid Rent for 2013 and 2014 but Did Not Provide the Required Certified Financial Statements 0.71
Airport Commission: Rylo Management, LLC, dba Three Twins Ice Cream, Reported Its Revenues and Paid Rent for 2014 and 2015 per Its Lease Agreement but Did Not Provide the Required Certified Financial Reports 0.56
Airport Commission: The Airport Improved Its Construction Project Oversight, but Change Management and Data Reliability Procedures Must Be Strengthened 0.57
Airport Commission: Transportation Network Company Operating Permit Audit – Lyft, Inc., Complied With All Tested Transportation Requirements but Underpaid the Airport $30,042 in Trip Fees 0.61
Airport Commission: Transportation Network Company Operating Permit Audit – Wingz Correctly Paid the Airport $32,964 for 8,562 Vehicle Trips in April - September 2015 and Complied With Most Transportation Requirements 0.61
All Ten Selected Organizations Complied With the San Francisco Administrative Code, Chapter 12G, by Not Using City Funds for Political Activity for Fiscal Year 2013-14 1.93
Bay Area Young Positives Did Not Maintain Adequate Support for Reimbursement Requests Submitted to the Department of Public Health 1.19
Total 1.35
VISUALIZATIONS
FIELDS
Search
qryRecDeptStatus_P...
AuditTitle
ClosedDate
ConcurID
DateLastResp...
DepartmentN...
DeptID
Fiscal Year
Issue Month
Issue Year
IssueDate
Length Open ...
Length Open ...
RecID
RecNumber
Recommenda...
Recommenda...
Values
Drag data fields here
FILTERS
Page level filters
Drag data fields here
Report level filters
Drag data fields here
DRILLTHROUGH
Keep all filters
Off
Drag drillthrough fields here
Main Dashboard
Overall Recommendations
Screencastify - Screen Video Recorder is sharing your screen.
Stop sharing
Hide
PAGE 2 OF 2
Type here to search
11:19 AM
8/23/2018

Qualitative Analysis Tools
(Atlas.ti)

Pros

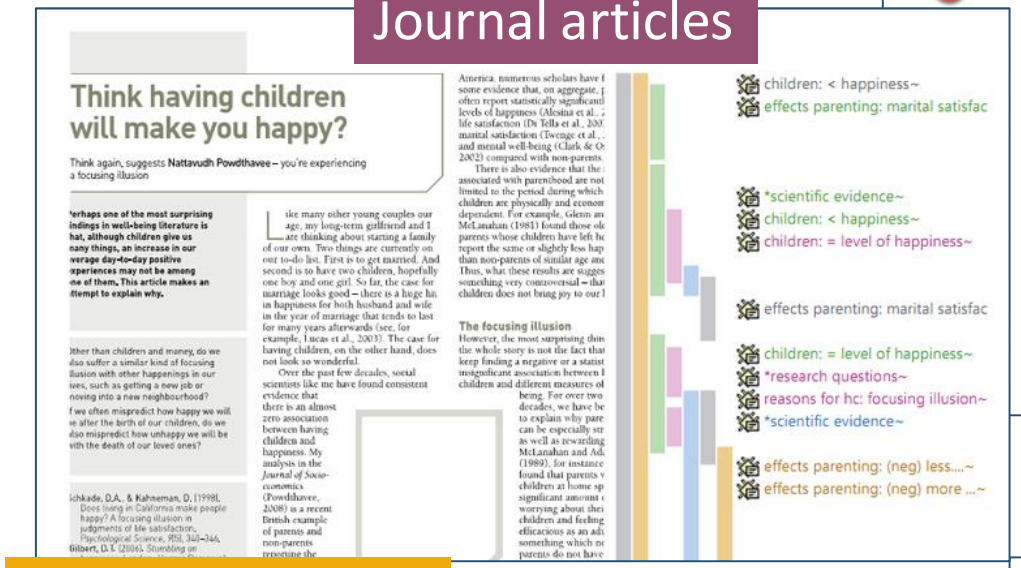
- Quantify qualitative data (best practice guidance literature, journal articles, interview records, audit reports, focus group and survey responses)
- Analyze a variety of media types including audio and video files
- Identify common themes across qualitative sources
- Identify statistically significant relationships between themes
- Turn anecdotal evidence into quantifiable evidence
- Excellent audit trail
- Excellent training materials and user guidance

Cons

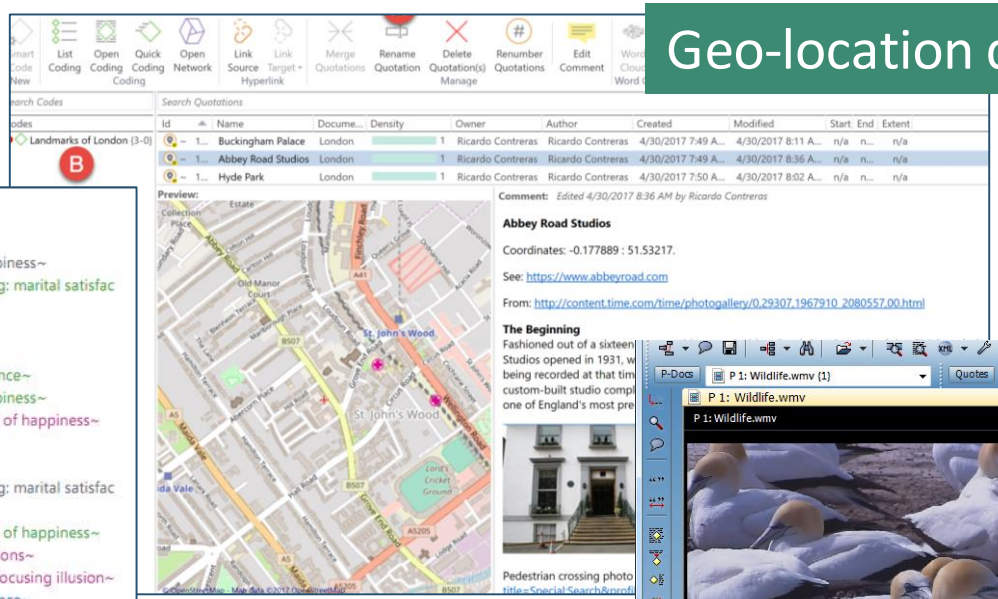
- Limited to qualitative data so will not be used on every project
- Requires a different mindset than other data analytics so may be a greater learning curve for audit shops with primarily accounting or business backgrounds

Atlas.ti

Journal articles

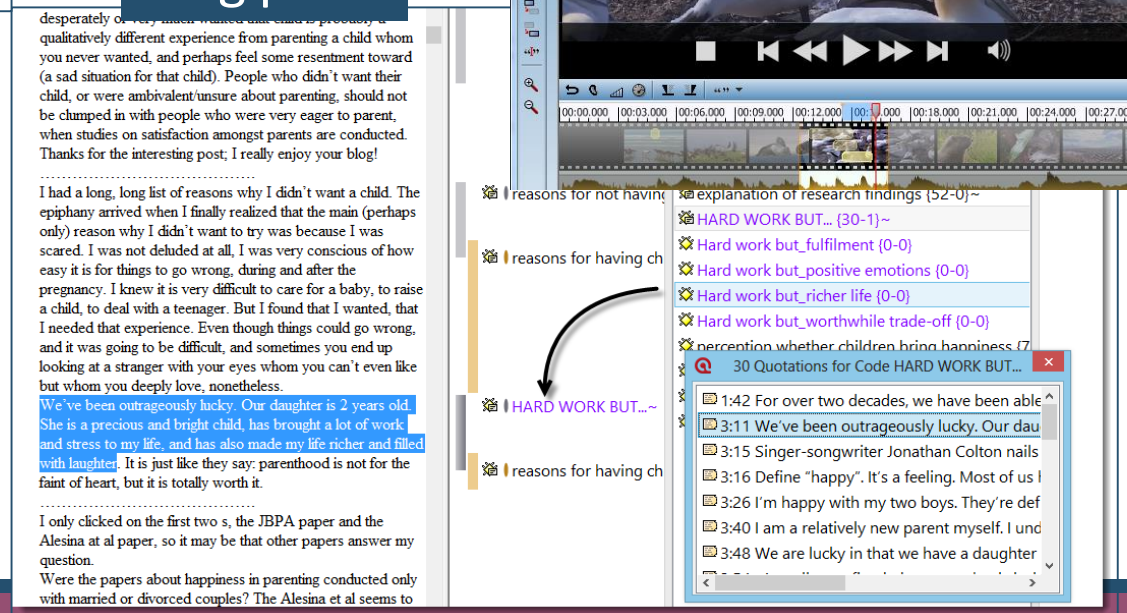


Geo-location data

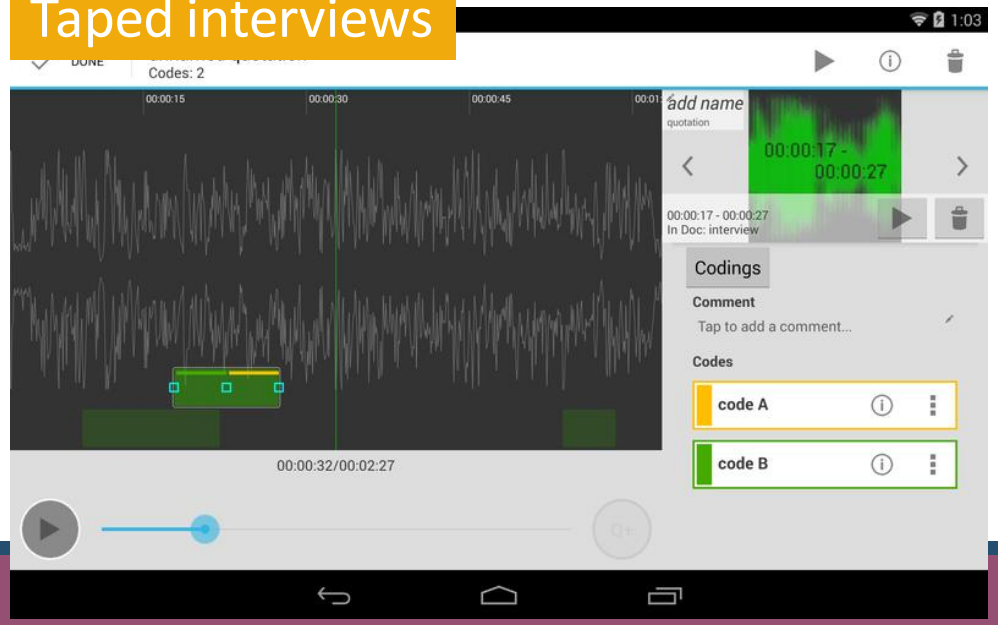


Video files

Blog posts



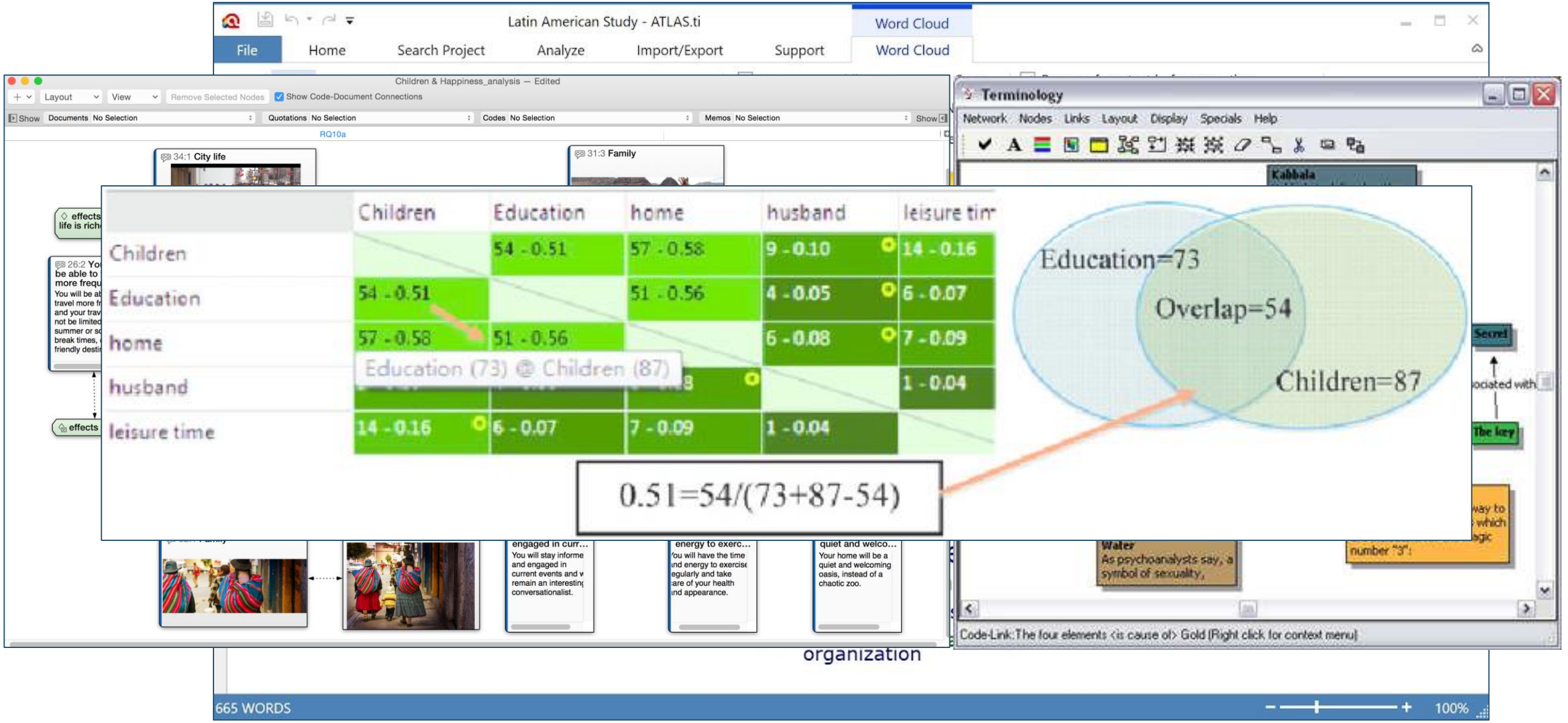
Taped interviews



Atlas.ti

Journal articles Blog posts Geo-location data Taped interviews Video files

Identify themes Map relationships between themes Identify co-concurrence of codes

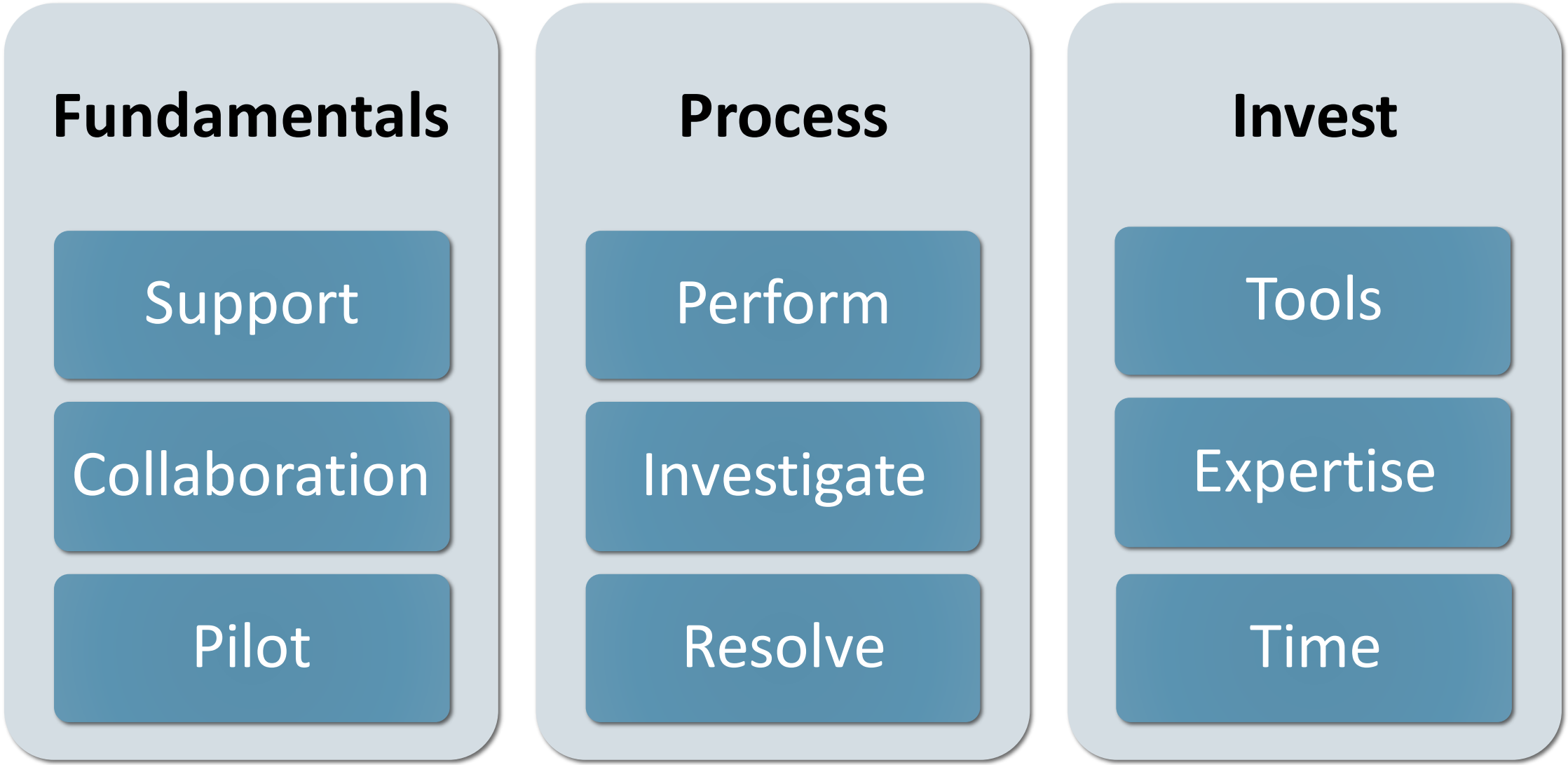


Data Analytics Tool Comparison

Tool	Cost	Granular Access Control	Capacity, Scalability, Performance	Data Analysis Functions	Statistical Functions	Visualizations	Ease of Use	Learning Curve	Audit Trail & Control	Server - based [1]	Interactive Output
Excel *(spreadsheet)*	Low	Low	Low	Mod	High	Mod	High	Low	Low	No	No
ACL *(audit software)*	Mod	Low	Mod	High	Low	Mod	Mod	Mod	High	No	No
Audit Exchange *(audit software)*	High	Mod	High	High	Low	Mod	Mod	Mod	High	Yes	No
TeamMate Analytics *(audit software)*	Mod	Mod	Mod	High	High	Mod	High	Low	High	No	No
Power BI *(visualization software)*	Low	Low	Mod	High	Low	High	Mod	Mod	Low	Mixed	Yes
Tableau *(visualization software)*	High	Mod	Mod	High	High	High	Mod	High	Mod	Mixed	Yes
SQL Server *(server database)*	Mod	High	High	Mod	High[2]	Mod[3]	Low	N/A[1]	Mod	Yes	No
Access *(desktop database)*	Low	Low	Mod	Mod	Low	Low	Mod	Mod	Low	No	No
Atlas.ti *(qualitative analysis)*	Mod	Mod	Mod	Mod	Low	Low	Mod	Mod	High	No	No

[1] Server based software requires IT staff to design and administrator
[2] SQL Server implements statistics using R statistical package
[3] SQL Server Reporting Services is used for reporting visualizations

Conclusion


Fundamentals
Support
Collaboration
Pilot
Process
Perform
Investigate
Resolve
Invest
Tools
Expertise
Time

Nelson Ho
Supervising IT Auditor
City and County of San Francisco
nelson.ho@sfgov.org

Kat Scoggin
Supervising Auditor
City and County of San Francisco
kathleen.scoggin@sfgov.org

Ron Yakus
Data Analysis and Strategic Applications Team Manager
Arizona Auditor General
ryakus@azauditor.gov

Bonus Content

Microsoft Excel – What you didn't know it could do

- Create relationships between tables like a relational database (Relationships)
- Import data from multiple tables in a relational database with greater ease than queries in Access (Power Pivot)
- Model changes and see output without altering the existing data (What-If Analysis)
- Automatically see forecasted values based on historical data trends (Forecast Sheet)
- Create and save perspectives to view different elements of the data
- Create hierarchies for splicing data such as year → quarter → month → day or country → state → congressional district (Power Pivot)
- Automatically identify invalid or illogical data fields (Data Validation)
- Collaborative editing (Office365)
- Statistics including regression, correlation, covariance, anova, and t-test (Data Analysis Toolkit Add-in)