

Tools for the Digital Audit

IDEA Data Analysis Software

PRESENTER:

Kevin Simon, Product Lead, Analytics



About CaseWare IDEA



- Wholly owned subsidiary of CaseWare International (Toronto, Canada) founded in 1988
- CaseWare IDEA Headquartered in Ottawa, Canada
- CaseWare IDEA brand established in 2015 to differentiate ourselves in evolving analytics marketplace
- CaseWare IDEA Offices in Netherlands, France, Belgium, Hong Kong and Costa Rica
- Over 35 Country Distribution Partners including Sama Audit Systems in India www.samaaudit.com

International Acceptance



CaseWare's Research & Development Team

CaseWare's R&D Teams are dedicated to validating and implementing new technologies that assist the auditor in:

- Gathering and organization of audit data
- Execution of the audit
- Presentation and use of audit results

Agenda

- Tools to Help Manage Data
- Tools to For Audit Analytics
- Other R&D Projects

Tools for Managing Data



CASEWARE®

A Well-Known Issue

- Getting and managing data is among the most time consuming aspects of any audit
- This area can be overlooked as 'plumbing'
- Much like real plumbing however, it's quite important, and problematic if not done correctly



<https://www.pinterest.ca/pin/478437160390917167>

CaseWare's Approach

- R&D Teams are developing tools to remove some of the burden
- As CSV is a common format for 'unknown' data sources, try to determine the meaning of each column based on the properties available to us:
 - Anything known about the source (Journal, Payroll etc)
 - NLP on column names if they exist
 - Type/Shape of data in each column

CaseWare's Approach: CSV Data Example

textsum.ipynb | column_tagging_row_model | data_services_intermediate_ | Python 3

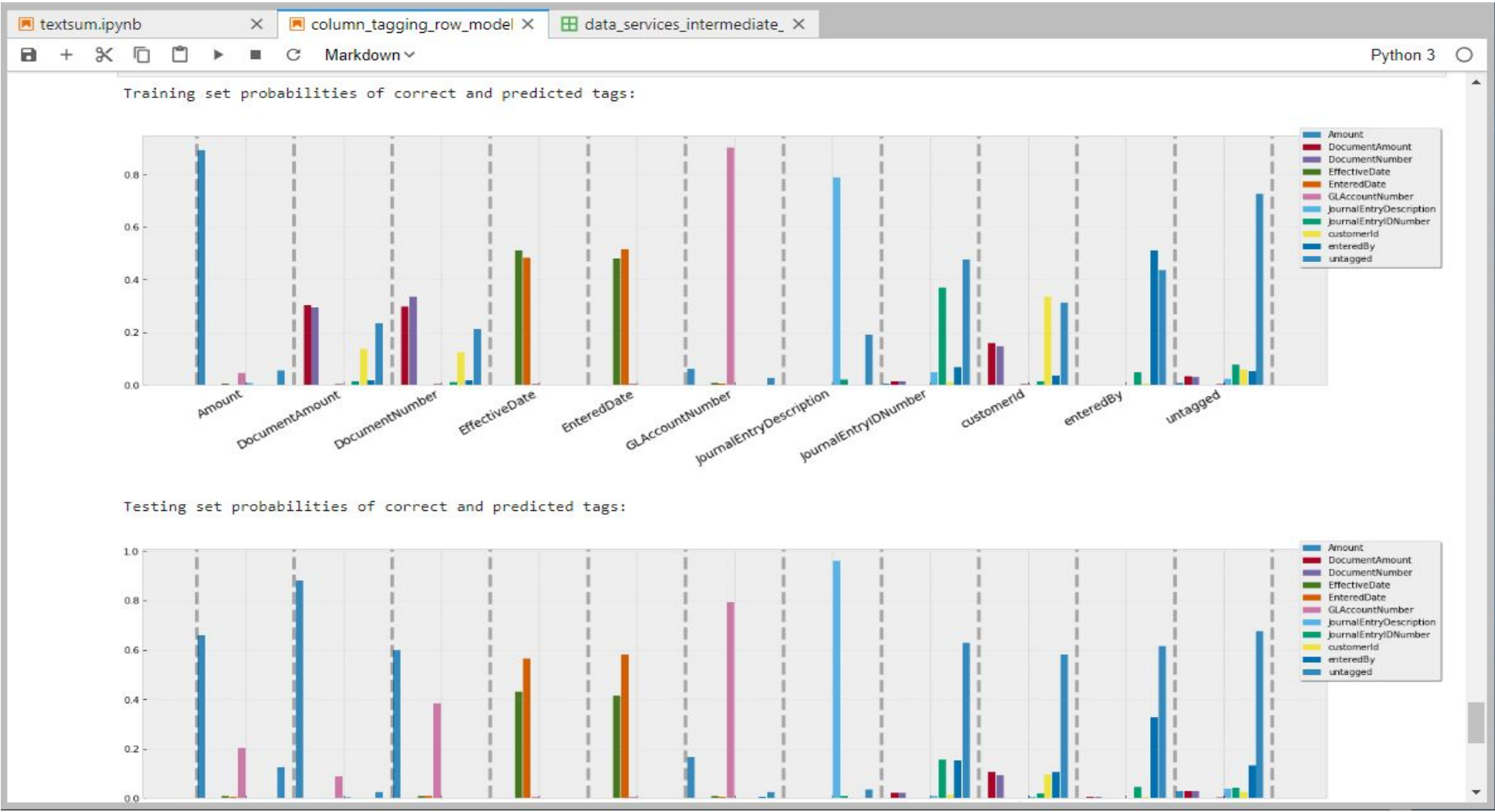
display first 10 rows of raw data

```
[5]: df.head(10)
```

	JE_NO	Acct_NO	Acct_Desc	Value	Trans_Date	Entered_Date	Entered_By	Trans_Desc	Ref_Type	Ref_NO	Invoice_Amnt	Customer_ID
0	SJ00001	45100	Consultancy Income	-122.73	2016-01-01	2016-01-21 06:44:00	Emma Flatt	Sale; VUO Reality	NaN	NaN	NaN	NaN
1	SJ00001	11700	Trade Debtors	12.27	2016-01-01	2016-01-21 06:44:00	Emma Flatt	Sale; VUO Reality	NaN	NaN	NaN	NaN
2	SJ00001	23050	Sales Tax Payable	-12.27	2016-01-01	2016-01-21 06:44:00	Emma Flatt	Sale; VUO Reality	NaN	NaN	NaN	NaN
3	SJ00002	11700	Trade Debtors	15.95	2016-01-03	2016-01-21 06:44:00	Emma Flatt	Sale; VUO Reality	NaN	NaN	NaN	NaN
4	SJ00002	23050	Sales Tax Payable	-15.95	2016-01-03	2016-01-21 06:44:00	Emma Flatt	Sale; VUO Reality	NaN	NaN	NaN	NaN
5	SJ00002	45100	Consultancy Income	-159.55	2016-01-03	2016-01-21 06:44:00	Emma Flatt	Sale; VUO Reality	NaN	NaN	NaN	NaN
6	SJ00003	11700	Trade Debtors	15.95	2016-01-10	2016-01-21 03:44:00	Emma Flatt	Sale; VUO Reality	NaN	NaN	NaN	NaN
7	SJ00003	23050	Sales Tax Payable	-15.95	2016-01-10	2016-01-21 03:44:00	Emma Flatt	Sale; VUO Reality	NaN	NaN	NaN	NaN
8	SJ00003	45100	Consultancy Income	-159.55	2016-01-10	2016-01-21 03:44:00	Emma Flatt	Sale; VUO Reality	NaN	NaN	NaN	NaN
9	SJ00005	11700	Trade Debtors	19.99	2016-01-28	2016-01-21 03:44:00	Emma Flatt	Sale; The Stationery Company	NaN	NaN	NaN	NaN

apply known maps from column name to Analytics tag ↕

CaseWare's Approach: After Some Proprietary Math



CaseWare's Approach: Prediction Results

```
textsum.ipynb | column_tagging_row_model | data_services_intermediate_ | Python 3
```

predict most likely tag for each column

Most likely based on most common most-likely tag per cell:

```
[28]: print("training (actual -> most likely)")
train_preds.groupby("actual").apply(lambda grp: Counter(grp["predicted"]).most_common(2)[0][0])

training (actual -> most likely)
[28]: actual
Amount Amount
DocumentAmount DocumentAmount
DocumentNumber DocumentNumber
EffectiveDate EffectiveDate
EnteredDate EnteredDate
GLAccountNumber GLAccountNumber
JournalEntryDescription JournalEntryDescription
JournalEntryIDNumber untagged
customerId customerId
enteredBy enteredBy
untagged untagged
dtype: object

[29]: print("testing (actual -> most likely)")
test_preds.groupby("actual").apply(lambda grp: Counter(grp["predicted"]).most_common(2)[0][0])

testing (actual -> most likely)
[29]: actual
Amount Amount
DocumentAmount Amount
DocumentNumber Amount
EffectiveDate EnteredDate
EnteredDate EnteredDate
GLAccountNumber GLAccountNumber
JournalEntryDescription JournalEntryDescription
JournalEntryIDNumber untagged
customerId untagged
enteredBy untagged
untagged untagged
dtype: object
```

CaseWare's Approach

- Trial balance account mapping
 - CSV, or in a known data format
- Known formats may require mapping to roll up in a consistent way
- Similar problem to CSV column tagging
- Similar approach to minimize the work the auditor needs to do

How Will We Learn - Data Lake

CaseWare's Unique Value:

- 30 years of customers using our products
- Reasonably consistent data model
- Ongoing good relations
- Millions of potential audits to mine for model building
- Ability for smaller firms to opt-in and gain the benefits currently availability to the Big 4

Data Management Tools

- Management of the data is key before any analytics can be done
- CaseWare is actively working on tools to help the auditor here
- Significantly less glamorous than some other R&D areas, but huge potential to streamline the audit workflow
- Additional research focuses on other potential tools and technologies

Tools for Audit Analytics



CASEWARE®

The Classic

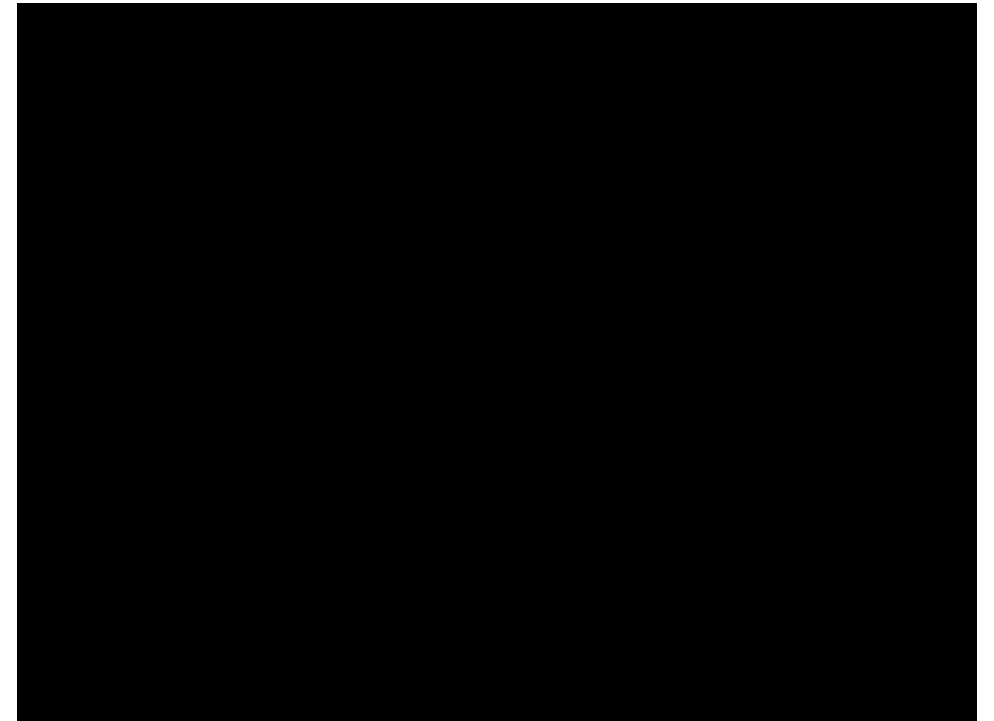


10.4 Release Date: November 19, 2018

- Better Python Integration

2019

- CaseWare Cloud Integration!



More IDEA

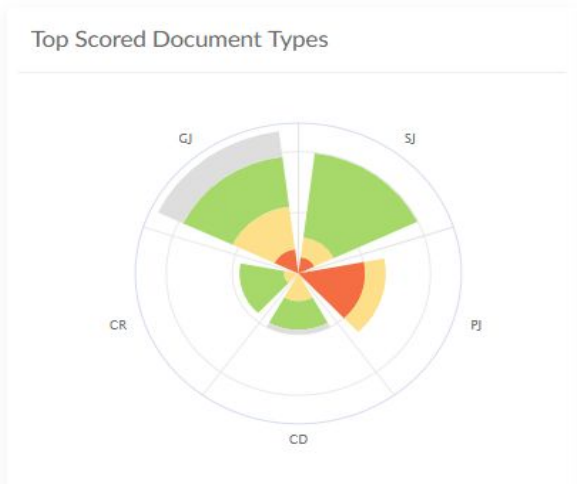
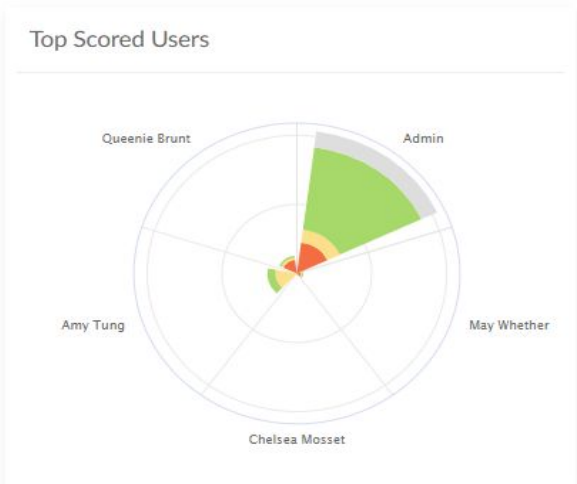
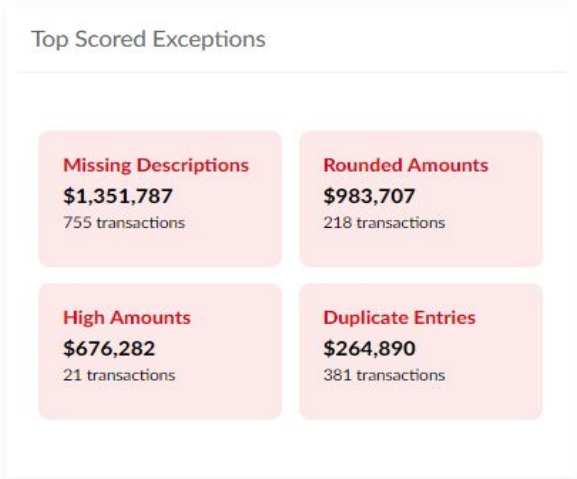
Natural Language Processing

- Opens up non-traditional data sources to the auditor
- Leverages Python integration

Outliers

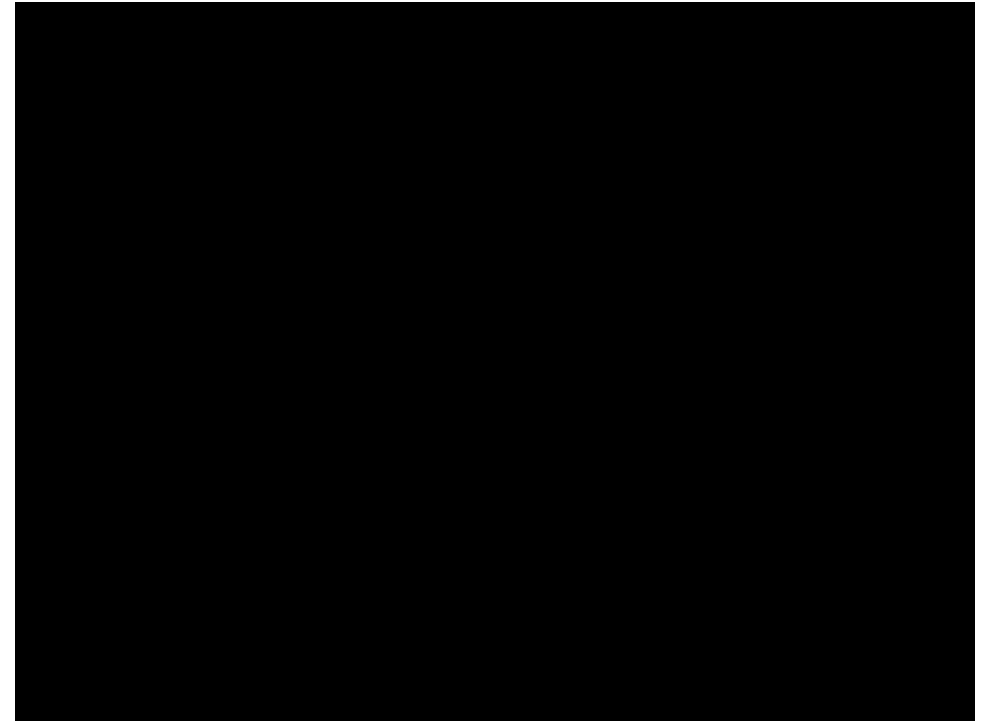
- Beta now available to download
- [http://ideascripting.com/IDEA Lab/outliers-detection-using-machine-learning-python](http://ideascripting.com/IDEA_Lab/outliers-detection-using-machine-learning-python)

Analytics in the Cloud



Process Mining

- Well known research topic
- CaseWare Approach:
 - Map the expected flows
 - Visually call out the exceptions



Analytics Cannot Replace the Auditor

- The auditor is still the key
 - Both the data and the automated processes needs to be understood
- Analytics augments the auditor for Efficiency and Accuracy

Other R&D Projects



CASEWARE®

Prediction of Misstatement Requiring Re-filing

Machine Learning, NLP, Ensemble Predictive Engine

ai.caseware.com

Synchronoss Technologies

<https://ai.caseware.com/company/0001131554>

<https://www.last10k.com/sec-filings/sncr>

Time Series Analytics

- Various techniques to examine data sets for cycles and trends
 - Cycles can be normalized out
 - Trends can be emphasized and examined
- Heavy use of Autocorrelation
- Fourier Analysis
- Techniques that are common in other types of analysis being applied to accounting data
- Can help find related transactions vs. single transactions

Thank You

QUESTIONS?

Contact:

- Kevin Simon: kevin.simon@caseware.com
- Visit: www.casewareanalytics.com



CASEWARE®

