Discover and Leverage the Gems in Your Data

Insights from a webcast sponsored by SAS and Casino Journal

Originally broadcast in May 2011

Featuring:

Mary Osborne, Text Analytics Architect at SAS

Kelly A. McGuire, PhD, Executive Director of the Hospitality and Travel Global Practice at SAS

"The hotel was very decrepit and very old. In short, the word 'dump' fitted the place exactly."

"The smaller money slot machines are always 'broken.'"

"There are always waits for machines unless you go very early in the morning or very late at night."

"The rooms were filthy and the employees were rude and unhelpful."

"Surly employees, dirty rooms and bland buffet food."

This is the worst casino I have ever been to in my life!"

These are actual review comments out there for all to see. It only takes a five-second Google search on "worst casino reviews" to find them. Whether such comments are posted on a travel review site, on Facebook, or on a travel writer's blog – and whether they're about your property or the competition – you'd want to know.

When you know what has dismayed or delighted guests, you have powerful insight into how to continually improve the guest experience – or how to capitalize on the competition's shortcomings. You get honest answers to the perennial questions for hospitality organizations:

- What are my guests saying about me?
- What are their most and least favorite things about my brand, hotel or casino?
- How do they feel about our brand versus our competitors?
- What kinds of issues are causing recurrent guest complaints?
- What new services or amenities do guests most want to see us provide?

The good news is that you have more ways than ever to get those answers. Guests express themselves not only in your surveys and call center communications, but also in a variety of customer-generated avenues, such as emails, online travel reviews, social media posts and online comments to others' blogs and posts.

Trouble is, there is almost too much of this information. A typical casino contact center, for example, might generate 70,000 call center records in a single month. What human would want to read all those records? What human could make sense of it all in an objective and systematic way?

"With the increasing speed of social media and the Internet in general, there are more and more ways we communicate with our guests, and we're creating unprecedented volumes of unstructured text data," said Kelly McGuire, Executive Director of the Hospitality and Travel Global Practice at SAS. "This is an emerging and critical data source."

"With the increasing speed of social media and the Internet in general, there are more and more ways we communicate with our guests, and we're creating unprecedented volumes of unstructured text data. This is an emerging and critical data source."

Kelly McGuire
Executive Director of the Hospitality and Travel Global Practice at SAS

How can the hospitality and gaming industry get a handle on all this valuable but chaotic text data and use it to generate meaningful insights and better decisions? That was the topic of a May 2011 webcast co-hosted by SAS and *Casino Journal*. In the one-hour, interactive event:

- McGuire described some practical ways to incorporate unstructured text data into business analysis.
- Mary Osborne, a Text Analytics Architect at SAS, explained how text analytics works in classifying content and revealing hidden patterns.

"Obviously one use of this data is to monitor the buzz that's out there and address the tactical, day-to-day comments and concerns" such as cold buffet food or a soiled room, said McGuire. "But when it comes to opportunities to solve business problems, there is a lot more out there than just this monitor-and-respond aspect, when you fold some of this unstructured text data into your routine analysis as an additional data source." For example, text data can augment traditional analysis in four key areas:

- **Brand management** – How is your brand performing against your competitors? As an enterprise, are you living up to the brand promise in the marketplace? Is each property measuring up? "You can use unstructured text data to understand how people are talking about your brand, and then compare these findings with whatever else you're doing to monitor the health of your brand," said McGuire.
- **Guest services** – "There are huge opportunities to use unstructured text data to augment the kinds of analysis you already do in guest services," said McGuire. "You can uncover service preferences to redesign your service offerings to better meet guest needs – or perhaps there's an element that's missing. Are guests saying they wish there was a better breakfast option available, or that your restaurants were open later at night? You can also uncover key areas of complaints: where you might need to do a renovation, improve a product or provide some additional training."
- **Public relations** – By monitoring social media and other ongoing communications in a consistent and automated way, you can determine the impact of a press release, see what kind of reaction it has created, and help to predict the impact of future public relations and marketing efforts.
- **Market research** – "Unstructured text data can be a huge component of a market research program," said McGuire. "In addition to the controlled, structured and formal analysis you already do through guest surveys and focus groups, you can bring in unstructured text data from internal or external sources to identify areas where competitors' offerings exceed yours, and see how you can change your offering to meet or exceed the competition."

If the webcast audience was representative of the industry at large, much of this opportunity is still untapped. In an informal webcast survey, 25 percent of respondents said their organizations collect data from customer surveys and feedback mechanisms but don't do much with it. The remaining 75 percent have at least assigned someone to read, summarize and categorize the information – usually with highly manual processes, but are at least trying to start doing something with it.

> "Text analytics enables you to evaluate the content in online conversations. Because these conversations are public, you can analyze review of your competitors and compare them against yours. This is a phenomenal opportunity that the industry has really never had before."
>
> **Kelly McGuire**
> *Executive Director of the Hospitality and Travel Global Practice at SAS*

However, you could have an army of people reading and interpreting all the unstructured text data around your brand, and it still wouldn't be enough. For one, humans are unwitting filters of information. Two people can read the same words and interpret them quite differently. For example, the comment, "The lobby bar was quiet" might be either: (A) a positive statement for a patron seeking a quiet place to confer with a colleague, or (B) a disappointment to a guest looking to mingle and socialize. Which is it, positive or negative?

Even if you could get all your readers to agree on every rule of interpretation, you couldn't get them to correlate their findings, for instance, to categorize 70,000 call center records, identify the words that appear most frequently, and see how those results align with customer survey results. Given the massive scale of text content available to hotels and casinos, only with automation can you really uncover and exploit the gems hidden within it.

An Inside Look at Text Analytics

When you think about automating the analysis of unstructured text, it's easy to foresee the challenges. "Unstructured text data is messy, it's long, it's complex, it moves faster every day, and we collect more of it every day," said McGuire. The sheer volume of data calls for powerful analysis – not to mention the need to somehow replicate the logic that humans use to understand language, in all its variations and nuances.

"Humans can dissect a sentence and get the main ideas out of it," said Osborne. "Machines have a hard time parsing language, because language is constantly evolving and devolving. For example, look at how people are abusing language when they text. Sometimes when I see texts come across my phone, I feel old because it seems like people are speaking a completely different language."

Here is where natural language processing (NLP) comes in. Natural language processing combines computer science and linguistics to identify meaningful concepts and attributes in spoken or written word, typically electronic documents, files or records.

Osborne explained some of the key facets of this linguistic science, which enables machines to do a respectable job extracting meaning and sentiment from unstructured text data, with some help from humans.

- **Part-of-speech (PoS) analysis** assesses the structure of the sentence to determine the part of speech – noun, adjective, adverb, etc. – such as the difference between "one cashier can service 30 tabs per hour," or "the cashiers have a new service protocol."
- **Style conventions** indicate the start and end of a word or sentence. Rules are unique for each language. For example, the beginning of a sentence in English is marked by a capital letter, but in German, every noun is capitalized. In English, words are separated by spaces and/or punctuation, but not in Chinese. These details are coded in rules.

- **Built-in lists** store knowledge associated with: *entities* (recognizable people, places, organizations, currency, etc.); *word stems* (such as plural, past and future tense variants of a word); *synonyms* (different words that have the same meaning), spellings (which may or may not be auto-corrected, depending on the need); and *filtering* (overlooking extraneous words, such as conjunctions or words you have identified as meaningless, such as RT for "retweet" in a Twitter feed).
- **Tokenization** translates multiple words into the concept those words represent, such as identifying "facilities manager" as a title, "food court" as a place, and Wayne Newton as a person.
- The context of the sentence aids with **disambiguation**, such as differentiating between wait staff or wait time for a slot machine, between AC as an acronym for air conditioning or Atlantic City, or between Paris Hilton the heiress or a Hilton property in Paris.

Some of the earliest forms of natural language processing used machine learning algorithms, such as decision trees that produced hard if-then rules from structured fields of categorized data. Likelihood statistics and even probabilities were then used to assign weights to categories.

Modern text analytics software goes beyond simply counting and comparing words that have been structured into relational fields. It now understands context at a deeper level by assessing words in relation to each other. For example, text analytics can:

- Understand abstract concepts, such as "blue plate special," even though the phrase is not about blue plates.
- Extract facts, such as differentiating between "The waiter returned with the dish" and "the waiter returned the dish."
- Discern attitude polarity: whether the content expresses a positive, negative, mixed or neutral sentiment, even if positive and negative terms are used together. So if Jane Austen is writing those travel reviews, text analytics understands that "not altogether disagreeable" is not a negative, even though it contains the negative terms "not" and "disagreeable."

These capabilities, coupled with the higher processing power now available, have made text analytics with huge data volumes practical for real-world business problems. It's time for hotels and casinos to take advantage.

Text analytics is the use of computer software to annotate and extract information from electronic text sources – finding the key concepts, patterns and facts – and analyzing that information for business purposes.

Practical Applications for Text Analytics

SAS thinks of text analytics applications in two broad categories: managing information assets and discovering patterns in text data. Rather than being standalone applications, the technologies in these categories are often used in complementary ways.

Text Analytics for Improving Information Access

Information access applications make information searches far faster and more effective than manual methods, with text repositories linked together through consistent and systematically defined relationships. The key capabilities in this category are content categorization and ontology management.

With **content categorization**, the system scans through a piece of information and categorizes it based on terms it finds. For example, a content categorization program can quickly scan and classify the thousands of comments coming in from guest surveys each day, and identify which comments relate to a housekeeping issue, the room, the bed and so on. Once categorized, it is easy to direct the comments to the right person or department.

"Content categorization is based on a predefined taxonomy, a logical (usually hierarchical) organization of the data," said Osborne. "For a hotel, the top of the taxonomy might have the main idea for a hotel, and digging deeper, you may have the room, and then in the room, the bathroom, and within the bathroom, the towels, sink, shampoo and conditioner."

Ontology management links text repositories together through consistently and systematically defined relationships. "It is the ability to manage multiple taxonomies or multiple vocabularies in an organization," said Osborne. "For example, the marketing organization and sales organization may be talking about the same things but using different verbiage. The same is true of organizations that have grown through mergers and acquisitions. Through ontology management, we can standardize the vocabularies and account for those differences across functional areas or business units."

> "Taxonomies are living breathing things that evolve over time. The more you test them, the more data you apply to them, the better they become."
>
> **Mary Osborne**
> *Text Analytics Architect, SAS*

Text Analytics for Pattern Discovery

"Pattern discovery is where we get into the true analytical side of text analytics," said Osborne. Within this broader category are technologies for:

- **Sentiment analysis**, which automatically locates and extracts sentiment from online materials
- **Text mining**, which provides powerful ways to explore unstructured data to discover previously unknown concepts and patterns

Sentiment analysis is a form of natural language processing that looks at how people use words and phrases in context, and then assigns a sentiment – positive, negative or neutral – based on the words people use.

"It is incredibly useful to apply this technique to online reviews and comments," said McGuire. You can classify and categorize these sentiments, look at trends, and see significant differences in the way people speak either positively or negatively about you – and your competition. "There is just so much potential in this emerging area. It is a natural fit for the hospitality industry."

Granted, there are some challenges. "This is always going to be tricky because we're talking about nuances of language," said McGuire. Words mean different things in different contexts – a menu item can be "the bomb" (good) or it can "bomb" (bad). Sarcasm is also hard to discern, and trends in language change with time and cultural shifts. "Fortunately, algorithms learn over time, so we uncover some of these issues. Despite the challenges, there's some real power in being able to automatically assign sentiment to thousands of documents."

"Statistical models generate tonal keyword lists of words that we can say are positive, negative or neutral," said Osborne. "Words such as 'difficult,' 'dislike,' 'poor' or 'disappointing' are pretty straightforward. Others can be ambiguous, so it takes some human intervention to get the best results. The statistical model comes up with the list, but then you combine it with some rules."

You don't have to be a linguist to define rules about features, products or offerings. For example, a rule could say that a sentence where casino games are mentioned with a positive tonal keyword is a positive statement. The next rule could say that where casino games are mentioned, if there is a negative tonal keyword such as 'not' within a distance of two words, it's negative, unless it negates a negative. For instance, if a comment says 'This slot machine is not bad,' that's good.

"Rules can get very granular to get very good results," said Osborne. "Rules are typically reused from project to project, so it doesn't take a lot of manual intervention; you just borrow and steal from yourself to create better analyses each time around."

Hotel deals are many in Vegas with the current financial crisis in USA.
I received a very good deal through Allegiant Air which included air/hotel.
I travelled with my 22 year old daughter as a graduation gift.
The room faced the pool which made it a little noisy but quiet at a reasonable hour.
Check-in was brutally slow and this seems to be the norm in Vegas recently.
The room was a stripped down old room.
The carpets were filthy, the bathroom old and provided with minimal toiletries including poor quality towels and terrible shampoo/conditioner.
No hotel information brochure and a phone service that went to message instead of a person to help with questions.
Beds were comfortable, radio/clock difficult to operate.
Would not stay at this property again.

Figure 1. Red indicates tonal negative keywords, blue for positive, and green for a feature.
If you see in this visualization that a rule isn't firing properly, you can refine the rule.

"A text-based report shows us exactly why certain things are considered positive or negative," said Osborne. "If an overall document was ruled as negative, we can see which aspects were hitting as positives, which were hitting as negatives, and in cases of mixed polarity, why the information was ruled as negative."

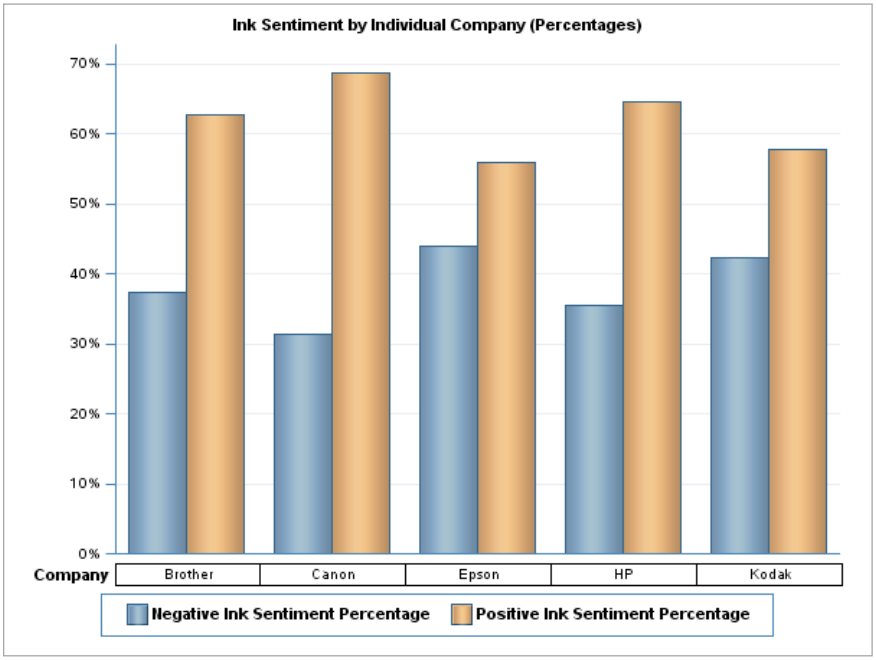

Figure 2. A visualization provides a quick view of positive and negative sentiment about different aspects of the business.

Text mining enables you to explore unstructured text data without a preconceived hypothesis, said Osborne. "When you're dealing with mountains of unstructured text data, and you're not really sure what's in that data, you can use text mining to give you a breakdown of some of the components that exist in that collection."

"You can dig around in volumes of unstructured data and look for themes, connected concepts and volumes of conversation," said McGuire. "You can see what your consumers are talking about most, and with what other concepts they're talking about. You start to uncover relationships among conversations and among linked concepts and see whether they are significant or not."

Terms

TERM	FREQ	# DOCS	KEEP ▼	WEIGHT	ROLE	ATTRIBUTE
rt	717	686	☑	0.157	Prop	Alpha
vegas	390	350	☑	0.246	Prop	Alpha
dior	231	230	☑	0.296	Prop	Alpha
⊞ lady	183	182	☑	0.326	Noun	Alpha
march	165	164	☑	0.34	Noun	Alpha
⊞ hotel	133	128	☑	0.374	Noun	Alpha
⊞ tote bag	115	115	☑	0.386	NOUN_GROUP	Alpha
⊞ model	114	114	☑	0.387	Verb	Alpha
⊞ girl	109	108	☑	0.394	Noun	Alpha
⊞ meet	111	108	☑	0.395	Verb	Alpha
⊞ casino	104	101	☑	0.404	Noun	Alpha
⊞ night	101	101	☑	0.402	Noun	Alpha
allkpop	93	93	☑	0.413	Noun	Alpha
allen	90	90	☑	0.417	Prop	Alpha
vance	90	90	☑	0.417	Prop	Alpha
importance	89	89	☑	0.419	Noun	Alpha
experience	89	89	☑	0.419	Prop	Alpha
black	84	84	☑	0.426	Noun	Alpha
black	84	82	☑	0.431	Adj	Alpha
⊞ key	81	81	☑	0.431	Noun	Alpha
⊞ issue	80	79	☑	0.435	Noun	Alpha
⊟ big	79	79	☑	0.434	Adj	Alpha
bigger	37	37			Adj	
biggest	6	6			Adj	
big	36	36			Adj	

Figure 3. A terms report shows you which words appear most frequently in the text data.

Having identified which terms are prevalent, concept linking shows how tightly correlated these terms are. The darker the line between words, the more highly correlated they are. "Concept linking can be helpful to bootstrap a taxonomy for content categorization," said Osborne. "The fact that these entries are correlated gives me some guidance as to how I might want to organize that data."

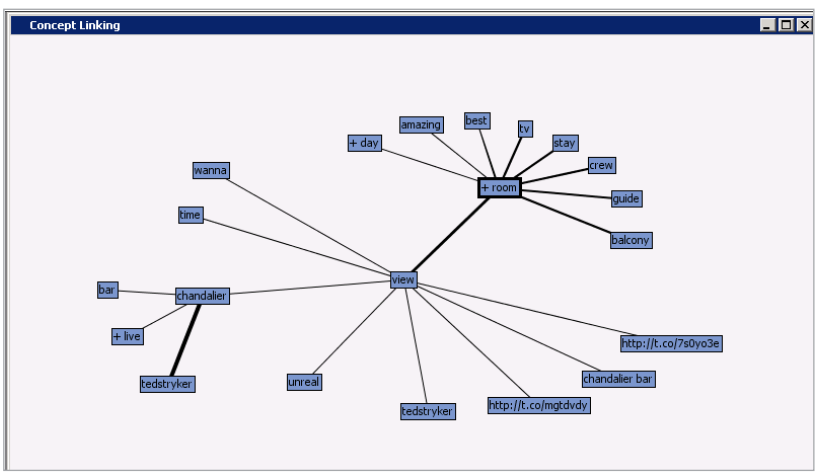

Figure 4. Concept linking shows relationships between the most frequently used terms in the text data.

Tips for Getting the Most from Text Analytics

Make it somebody's mission.

"It doesn't take an incredibly specific skill set to start taking advantage of text analytics," said Osborne. "If you're familiar with language, it's fairly easy to begin to write some of the rules we've been talking about. And if you have a background in analyzing structured data, it's easy to make the switch to text mining.

"Generally the people I see working in text analytics have worked in other areas, such as statistical analysis or business intelligence reporting. You don't have to be a linguist or have a PhD in statistics. It comes down to understanding the subject matter. If you have someone who has been with the organization for a while, who understands the voice of the customer, very often this person can hit the ground running with text analytics technologies and get some good answers."

Make sure you're working with relevant data.

The data collection methods will have a big influence on the quality of the raw data you start with, said Osborne. "You have active data collection methods, such as surveys, feedback forms or pulling in data from contact centers. Since these are all data collection methods that you initiate, you can guarantee that the majority of the data you're collecting will be relevant to your organization."

Not so with data collected from social media and other online sources through Web crawling of Twitter or Facebook feeds, travel blogs or discussion forums, for example. Your search could turn up a lot of noise. If the name of your company means "green" in Romanian, be prepared to get back a lot of Romanian documents. Fortunately, SAS® Text Analytics understands 30 languages and auto-detects the language being used, but if your organization doesn't have a presence in Romania, you're still going to get irrelevant results.

"You could very easily go out and do a very simple Twitter pull, analyze that data, and come out with a report that is complete garbage, because the majority of the data may have nothing to do with your organization," said Osborne. "It comes down to having business rules in place to define what is really relevant in the data. It could be as simple as writing very basic SQL query to pull out those records that are relevant, and evolve this process over time."

Consider a data pull for records that mention SAS. Rules could say that if the document included words associated with shoes or airlines, it probably relates to the other companies named SAS. But if the document includes words associated with analytics, it's a good bet it refers to SAS the analytics company.

"If you have someone who has been with the organization for a while, who understands the voice of the customer, very often this person can hit the ground running with text analytics technologies and get some good answers."

Mary Osborne
Text Analytics Architect, SAS

"It does take a little bit of looking and reading, a little bit of research to determine if the data is relevant, but it's not as daunting as it may sound," said Osborne.

"Another way you can eliminate a lot of noise is by knowing the business question that you're trying to answer, as opposed to just approaching the data saying, 'Let me see what's going on in social media,'" said McGuire. "Am I trying to investigate a guest satisfaction problem, a public relations approach, or reaction to my lobby redesign? If you consider the business problem first – then look at all the streams of data as pieces of the puzzle to solve that specific problem – you can weed out a lot of noise and a lot of the potential for erroneous analysis."

Consider demographic differences in language use.

A difficulty in automating text analytics is that language is constantly evolving and devolving, said Osborne. "To a machine, it can be challenging to determine whether people are actually saying something good or bad about your organization. For example, is something bad, or is it Michael Jackson's version of bad, which was actually good? Is 'sick' a good thing or a bad thing? It depends on the context and the age of the speaker. Is 'boss' really cool or a manager? It depends on how it is used in the sentence.

"We have to take those kinds of things into account when we're analyzing unstructured data. Then again, sometimes you can gain insights into demographics by the language being used, and at other times, you can apply what you know about demographics to figure out if the words are being used in a positive or negative light."

Use text analytics to discover your problem areas.

Is the call center at your casino doing a good job at handling inquiries, reservations, problems or special requests? If there are issues, you would want to know quickly, to improve the customer experience and determine where to put your focus.

"We did a project for a customer that thought they knew where they ought to focus," Osborne recalled. "When we ran their customer service logs through a taxonomy, to identify the main ideas there, we found that they were actually focusing on an area that wasn't very important."

The company had implemented scripts to improve consistency of call handling, but customers were turned off by agents who spoke like robots. Parroting from a script came across as cold and impersonal, and service quality declined whenever the agents had to deviate from the script. "Through content categorization and text mining, we were able to really get to the heart of the voice of the customer, and determine where they ought to focus."

> "Text analytics enables you to look at all of the information holistically, rather than just focusing on numbers or just the comments. You have to look at the entire puzzle."
>
> **Mary Osborne**
> *Text Analytics Architect, SAS*

Validate or clarify what structured data analysis tells you.

"Organizations typically collect survey data based on a numeric scale, and historically we have analyzed those surveys just by looking at the numbers," said Osborne. "The numbers don't tell the whole story, because in addition to the numeric values, you've got an unstructured component from people who are willing to give you extra feedback."

The written feedback might even mitigate a harsh numeric score. Survey respondents often feel bad about dishing out a low score and feel compelled to say something nice to balance it out. The guest may bash the front desk staff as a "1" but hasten to point out that the housekeeper was fantastic, wonderful and attentive.

"If you're just focusing on the numbers, you would only capture that negative aspect and you would miss the positive things that are going on with the housekeeping staff," Osborne said. "There are so many things you can miss out on if you only focus on the numbers. The good and bad experiences may even cancel each other out. Just because somebody had a singular bad experience with a front desk person doesn't mean they're not going to return. Text analytics enables you to look at all of the information holistically, rather than just focusing on numbers or just the comments. You have to look at the entire puzzle."

Closing Thoughts

As of March 2011, there were 175 million registered Twitter users, more than 100 million professionals on LinkedIn and 640 million Facebook users worldwide, half of whom log into Facebook each day, sharing more than 30 billion pieces of content a month. In the previous 12 months, Twitter and LinkedIn membership had more than doubled, Twitter traffic (tweets) soared 250 percent, and Facebook added another 200 million users.

"The volume of unstructured data coming from social media, even for a brand that is just marginally recognized, is estimated to be up to a terabyte a day," said Osborne. Add to that the tens of thousands of call center records, guest surveys and emails and other internal data sources, and you have a rich view into customer opinion – if you choose to use it.

Ignore it at your peril, said Osborne. "We've all seen situations where brands have been broken down by the voice of the customer. All it takes is a couple of bad experiences and it can spiral. There are Internet watchdogs out there who are constantly waiting for something bad to happen so they can blog about it. Text analytics gives you another mechanism to try to protect your brand. We want to make sure to stay ahead of that."

"We've all seen situations where brands have been broken down by the voice of the customer. All it takes is a couple of bad experiences and it can spiral. Text analytics gives you another mechanism to try to protect your brand."

Mary Osborne
Text Analytics Architect, SAS

About the Presenters

Mary Osborne

Text Analytics Architect at SAS

Mary Osborne is a business analytics specialist with more than 10 years of experience at SAS. Her expertise spans a variety of technologies, from data integration to business intelligence and analytics. Her primary focus is on text analytics, which includes text mining, content categorization and sentiment analysis.

She has a passion for using technology to solve business problems and feels that it's only through analytics that organizations can make good decisions. Osborne has worked in many different industries, including health and life sciences and retail. She enjoys helping her existing customers gain more value from their SAS investments and new customers find ways to improve productivity and profitability.

Kelly A. McGuire, PhD

Executive Director of the Hospitality and Travel Global Practice at SAS

Kelly McGuire works with product management, sales, alliances and R&D organizations to set strategic directions, drive product offerings, and ensure that SAS solutions meet the needs of the hospitality and travel market. She works closely with IDeaS Revenue Solutions, a SAS company, helping to integrate the IDeaS® revenue management solution with SAS marketing solutions.

She has 20 years of experience in the hospitality industry, both in operations and IT. Before joining SAS, McGuire consulted with Harrah's Entertainment, was a senior consultant at Radiant Systems, and worked for RMS (Restaurant Revenue Management Solutions). McGuire has a PhD from the Cornell School of Hotel Administration with a focus on nontraditional applications of revenue management.

For More Information

To view the on-demand webinar:

https://event.on24.com/eventRegistration/EventLobbyServlet?target=registration.jsp&eventid=299086&sessionid=1&key=91871C7B035B018475BD85B600E7FF1C&sourcepage=register

For more about analytics for the hospitality and gaming industry, visit SAS on the Web at www.sas.com/industry/hospitality/.

SAS Institute Inc. World Headquarters +1 919 677 8000

To contact your local SAS office, please visit: **www.sas.com/offices**