



(/)

## VANTAGE POINT (/VANTAGE-POINT-BLOG)

ADVERTISEMENT

# Data Hygiene: Big Data's Big Problem

By Dr. Cihan Cobanoglu, McKibbon Endowed Chair, Director of M3 Center & Coordinator of International Programs for CHTL - 09/03/2019



Data is one of the most important, strategic tools for any company, and the hospitality industry is no exception. It is now widely accepted that big data and its applications can be source of competitiveness. This competitiveness is driven by the

processes surrounding big data analytics. These processes include the collection, analysis and application of data according to the needs specific to each company and industry (Krsak & Kysela, 2016). Big data analytics is the application of advanced analytic techniques on big data sets (Russom, 2011).

Data mining is a process of big data analytics which is realized through four steps: data collection, data cleansing, data analysis and data interpretation (Krsak & Kysela, 2016). For a given company, the quality of data, which ultimately affects the reliability of analytics, depends on its fit with the specific needs for which it has been collected. In such conditions, the data cleansing step appears as a central and determinant one in big data analytics.

Data cleansing consists of checking for errors to make sure that the data collected is consistent and properly recorded (Krsak & Kysela, 2016). “Dirty data” - the term used for errors in data sets – is usually comprised of outdated data, incomplete data, or duplicate records. Big Data also needs to be cleansed either when it doesn't come from the same source system and needs to be integrated (i.e. in the case of mergers or acquisitions) or when it needs to be transformed in order to be harmonized (Watson, 2002). Thus, the cleansing of data is performed based upon specific rules related to the company or the industry. These processes, aimed at making the data “clean,” can be understood to be a part of “data hygiene.” Although it is a concept that is not much discussed in the literature, data hygiene is simply keeping the data clean to ensure that no duplicate, incomplete, outdated or corrupt data exist in the data sets (Kulshrestha, 2015).

The hospitality industry is considered the second largest industry in the world (WTTC, 2019) with 1.4 billion international tourist arrivals in 2018 (UNWTO, 2019). This high number of tourists represents a myriad of opportunities in terms of data generation and collection. Proper use of big data analytics could improve business decisions in the industry. Customer data in the hospitality industry can come from reservations, contact information, demographics, and also from social media (i.e. through reviews on sites such as TripAdvisor or Holiday Check) or from housekeeping. But as it is the case for each industry, difficulties may arise in the

data mining process, as many times collected data is not error free. Therefore, to achieve the most desirable outcome of big data analytics requires hospitality companies take the step of data cleansing.

Data hygiene can also help hotels maintain customer loyalty, operational efficiency as well as mapping their customer database (Kulshrestha, 2015). With the rise in misleading interpretations of data – due to dirty data sets – there has also been a rise in emphasis on better data cleansing techniques. Additionally, the sheer amount of the data needing to be cleansed suggests that automation of data cleansing would result in increased efficiency in the processes and outcomes of big data analytics. For instance, dailypoint™, a big data platform, has developed a data cleansing software specially designed for hotels that will allow for automatically cleaning, correcting and duplicating customer profiles (Leitsch, 2019). Such software makes the data cleansing process smoother, and likely allows for automatic harmonization of the data.

In the hospitality industry, marketing data could be used to target specific offers to certain customer segments (Krsak & Kysela, 2016) to reduce costs (revenue data). The influence such data could have on hotel revenues definitely requires this data to be clean in order to allow for optimal decisions in terms of marketing and revenue management. Today, there is some data cleansing software available that can help actors in the hospitality industry obtain the best insights from big data analytics.

## References

Krsak, B., & Kysela, K. (2016). The use of social media and Internet data-mining for the tourist industry. *Journal of Tourism and Hospitality*, 5(1).

Kulshrestha, S. (2015). Data Hygiene - Next Level Revenue Management for Hotels. Retrieved from <https://www.linkedin.com/pulse/data-hygiene-next-level-revenue-management-hotels-kulshrestha/>

Leitsch, H. (2019). [dailypoint™ Announces Data Cleansing Technology Partnership with protel](https://www.hospitalitynet.org/news/4092055.html). Retrieved from <https://www.hospitalitynet.org/news/4092055.html>

Russom, P. (2011). Big Data Analytics. Towi Research

UNWTO. (2019). International Tourist Arrivals Reach 1.4 billion Two Years Ahead of Forecasts. Retrieved from <http://www2.unwto.org/press-release/2019-01-21/international-tourist-arrivals-reach-14-billion-two-years-ahead-forecasts>

Watson, H. J. (2002). Recent developments in data warehousing. *Communications of the Association for Information Systems*, 8(1), 1.

WTTC. 2019. Travel & Tourism continues strong growth above global GDP. Retrieved from

<https://www.wttc.org/about/media-centre/press-releases/press-releases/2019/travel-tourism-continues-strong-growth-above-global-gdp/>

[DATA & ANALYTICS \(/DATA-ANALYTICS\)](#)

[HOTELS \(/HOTELS\)](#)

## **MORE BLOG POSTS IN THIS SERIES**

---



(/how-pieology-scores-four-star-service-rating)

## **HOW PIEOLOGY SCORES A FOUR-STAR SERVICE RATING**

(/how-pieology-scores-four-star-service-rating)

Wi-Fi marketing platform was the secret ingredient to improving customer satisfaction



(/how-solve-restaurant-industrys-technology-failures)

## HOW TO SOLVE THE RESTAURANT INDUSTRY'S TECHNOLOGY FAILURES

(/how-solve-restaurant-industrys-technology-failures)

What will it take to move the restaurant industry forward with meaningful technology standards? Jordan Thaeler, President and Co-Founder, WhatsBusy, explains.



(/facing-financial-management-challenges)

## **FACING FINANCIAL MANAGEMENT CHALLENGES**

(/facing-financial-management-challenges)

Laird Management Controller shares how the Burger King franchisee worked to replace manual, tedious practices with a scalable financial management system.




---

## BLOG AUTHOR

### GUEST VANTAGE POINT AUTHOR (/AUTHOR-PROFILE/VANTAGEPOINT)

Vantage Point is a recurring blog by guest authors. Contact us for more info on becoming a Vantage Point author. *[Read More \(/author-profile/vantagepoint\)](/author-profile/vantagepoint)*.

---

Send a Message  
 (/contact/profile\_contact\_form?blog=13774)

---

## SHARE



(<https://www.facebook.com/sharer/sharer.php?u=https://hospitalitytech.com/data-hygiene-big-datas-big-problem>)



(<https://twitter.com/home?status=https://hospitalitytech.com/data-hygiene-big-datas-big-problem>)



(<https://www.linkedin.com/shareArticle?mini=true&url=https://hospitalitytech.com/data-hygiene-big-datas-big-problem&title=Data Hygiene: Big Data's Big Problem>)



(mailto:?subject=Data Hygiene: Big Data's Big



Problem&body=I%20wanted%20to%20share%20this%20page%20with%  
hygiene-big-datas-big-problem)

---

ADVERTISEMENT

---

## YOU MAY ALSO LIKE



(/how-iot-could-  
take-hospitality-  
another-level)

INTERNET OF THINGS (IOT)  
(/INTERNET-THINGS-IOT)  
How IoT Could Take  
Hospitality to Another  
Level (/how-iot-could-  
take-hospitality-  
another-level)

---

## RELATED TOPICS

**DATA & ANALYTICS** (/DATA-ANALYTICS)

**HOTELS** (/HOTELS)

## GET THE NEWSLETTER



## **Subscribe**

**([https://ensembleiq.dragonforms.com/loading.do?pk=brandsitenlet19&omedasite=HT\\_prefnew](https://ensembleiq.dragonforms.com/loading.do?pk=brandsitenlet19&omedasite=HT_prefnew))**

## **GET THE MAGAZINE**

**Subscribe ([https://ensembleiq.dragonforms.com/HT\\_land](https://ensembleiq.dragonforms.com/HT_land))**

**Contact Us (</contact-us>)**

**About Us (</about>)**

**Media Kit (<https://fs6.formsite.com/stagnito/form143/index.html>)**

**(<http://www.ensembleiq.com/home>)**

© 2019 EnsembleIQ, All Rights Reserved