

# A SIMPLIFIED RESTAURANT LOCATION ANALYSIS USING MACHINE LEARNING



## CASE STUDY COURSERA CAPSTONE PROJECT

Francesco di Cugno

# Select a location in Munich for an Italian fine dining service

- There are a number of factors to be considered to identify a suitable **location** for restaurants

Culinary Trends	Demographic	Neighborhood	Infrastructure	Competition
<ul style="list-style-type: none"><li>• Is the offer in line with the consumer trends?</li></ul>	<ul style="list-style-type: none"><li>• Is there a customer base that is willing to buy the service and pay a premium for it?</li><li>• Is there a chance for recurring clients?</li></ul>	<ul style="list-style-type: none"><li>• Which venues are available nearby?</li></ul>	<ul style="list-style-type: none"><li>• Is the restaurant reachable?</li><li>• Is there any public transportation connection?</li><li>• Are there parking areas?</li></ul>	<ul style="list-style-type: none"><li>• How many restaurants/dining services are nearby?</li><li>• Are they premium ones?</li></ul>

Analyze **success factors** to identify evidences that will support recommendations for the selection of an **appropriate location in Munich** for an **Italian fine dining service**

# Data collection and preparation

## Demographic data

- Boroughs data was collected merging the following sources:
- (Wikipedia, 2018): containing the list of the boroughs, the number of residents, and the surface;
- (Statistisches Amt München, 2019): containing the number of residents in the boroughs, their age group (<6, 6-14, 15-44, 45-64, >=65), and the number of households per number of households (1 to 5);
- (TZ, 2011): containing the pre-capita income in the boroughs
- Result of preparation: 25 boroughs. 1 missing value for pre-capita income replaced by average.

## Nearby venue

- Foursquare used as data provider
- Selected the top 100 venues in the neighbourhoods within a radius of 4km from the neighbourhood centroid
- Result of preparation: 1457 venues found in the 74 neighbourhoods grouped by category

## Neighbourhoods geographical data

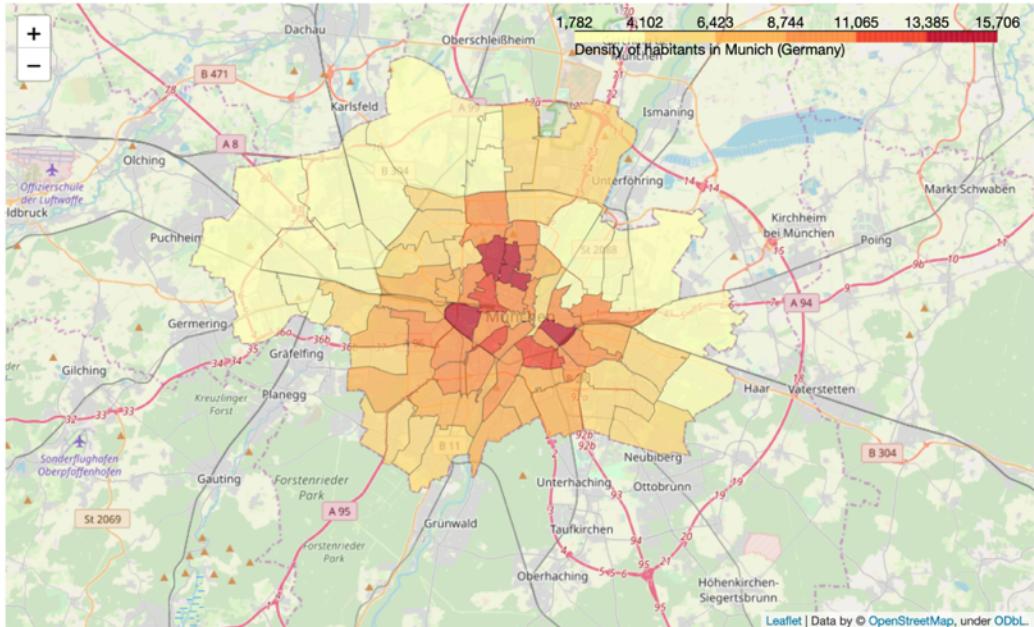
- Landeshauptstadt München, 2019): containing the list of the postcode of the neighborhoods belonging to the boroughs in Munich;
- (SUCHE-POSTLEITZAHL.ORG, 2019): containing the centroids of the neighborhoods in Germany and their surface.
- Result of preparation: 74 neighborhoods, 44 of which belonging to different boroughs

## Selected Features

Category	Feature	Reason for selection
Venues	1_MFV, 2_MFV, 3_MFV, 4_MFV, 5_MFV, 6_MFV, 7_MFV, 8_MFV, 9_MFV, 10_MFV	The top 10 most frequent venues (MFV) provide intrinsic information about resident preferences
Demographics	Density, Foreigners, Income, Age_28_44, Age_45_64, 1HH, 2HH_Adults, 3HH_Adults, 4HH_Adults, 5HH_Adults	The features support the selection of the neighborhoods considering the target customer base.
Pre-Capita Income	Income	The feature is an indicator of the buying power of the residents in the area
Venues	Count of restaurants, bar, and similar services in the neighborhoods by category	The list of all the restaurants and similar services in the area can be used to estimate the degree of rivalry in the neighborhoods.

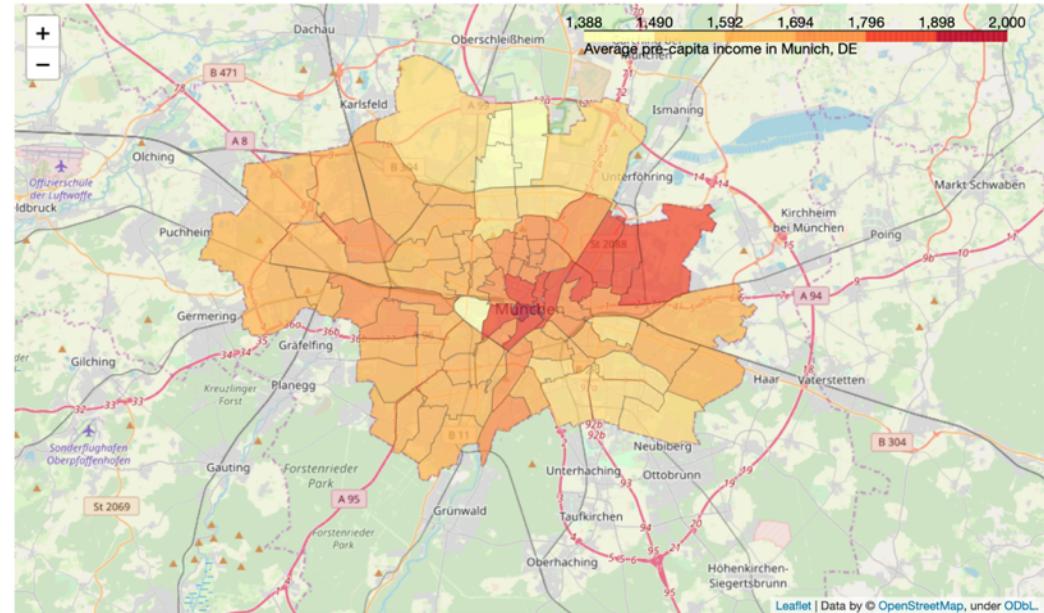
# Demographic data exploration

## Demographic data



- The distribution of the residents density has a standard deviation of 4093.98, indicating values spread out over a wide range.
- Neighborhoods in the center and south of the city have a higher density of residents compared to the peripheral ones.

## Pre-capita income

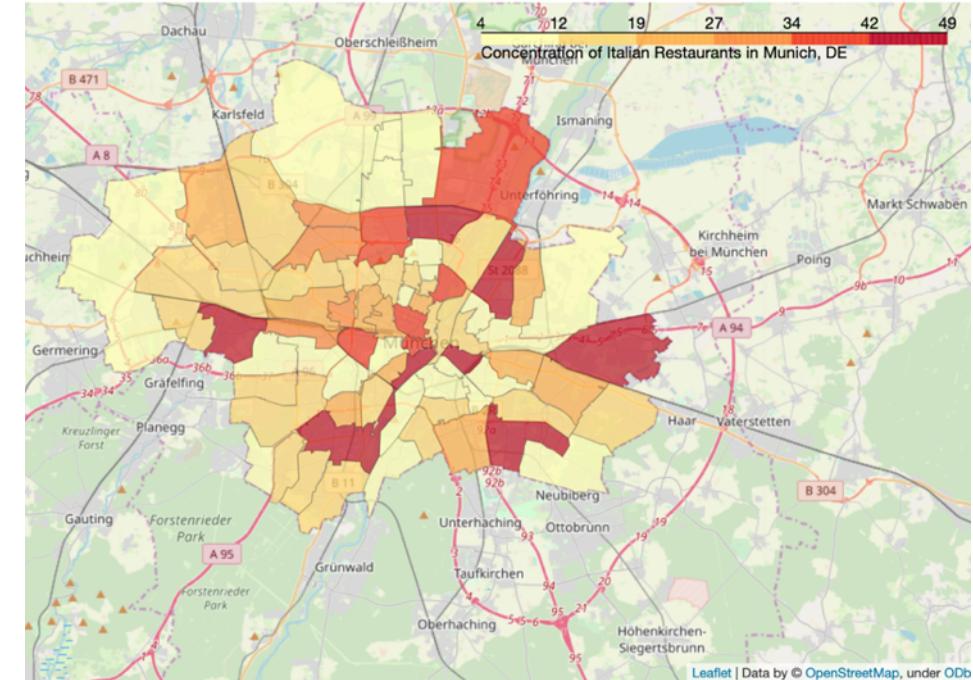


- The per-capita income has a standard deviation of 132.82, indicating values concentrated around the mean
- Also developing areas show income in the mean range
- Only Feldmoching-Hasenbergl (80933), Schwanthalerhöhe (80339), Milbertshofen-Am Hart (80937) have pre-capita income within the 25th quantile

# Location data exploration



*'Italian Restaurant'* is the second most frequent venue category in Munich



- The concentration of restaurants does not correlate with the resident density. Indeed, Trudering-Riem (81829) and Bogenhausen (81925) report a density within the lowest 25th percentile but are within the upper percentile for number of Italian restaurants.
- Normalizing the values by the number of habitants in the age of interest (adults from 28 to 64 years old), only Altstadt-Lehel (80331) remains in the upper 75<sup>th</sup> percentile, confirming the market saturation in the area.

# Neighbourhoods segmentation

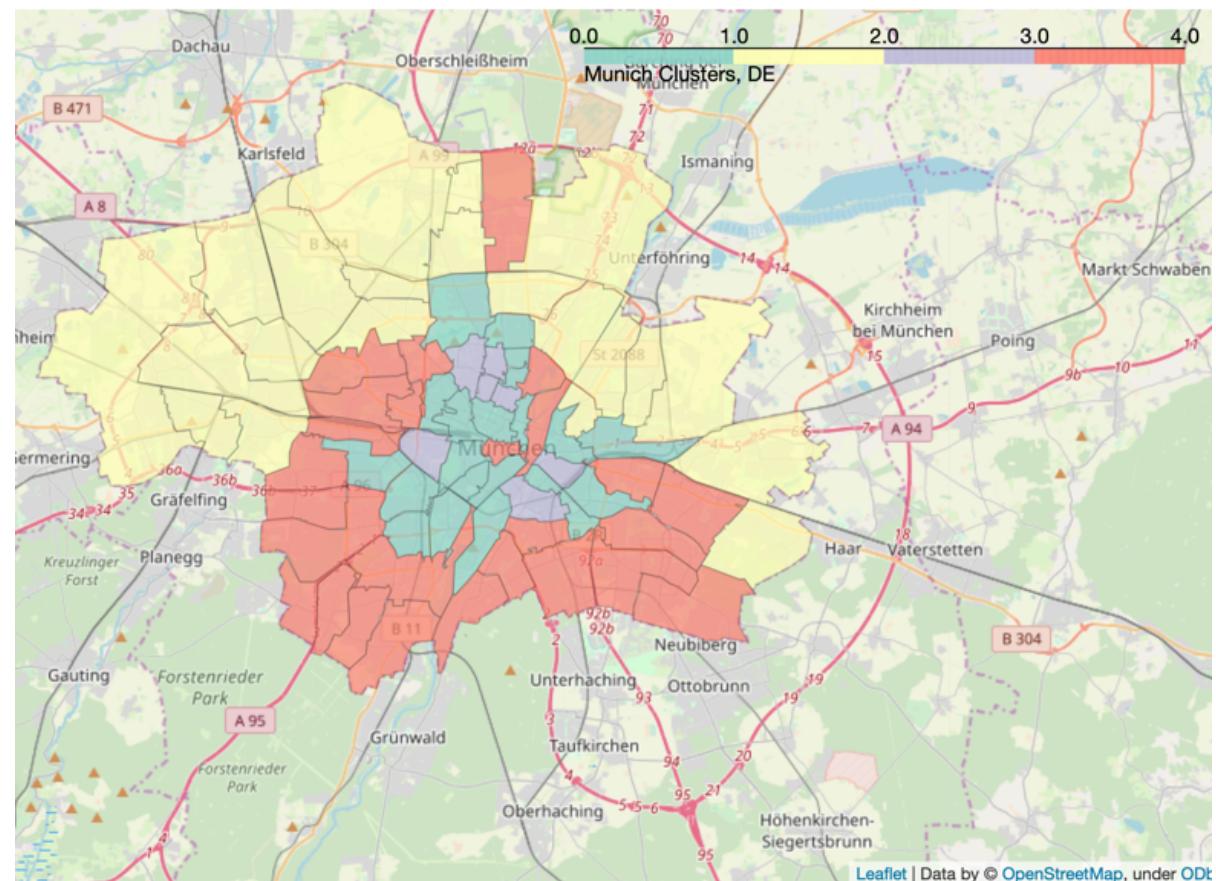
- The **k-means clustering** algorithm was used to partition neighborhoods in **4 clusters\*** to identify patterns that can support recommendations

**0**  
*High income. Number of 1 household larger than any other cluster. Very high resident density.*  
This cluster is characterized by high income and high concentration of 1 household. The density of residents is also very high.

**1**  
*Mid income. High concentration of families and foreigners. Fast food and dining services exceed restaurants.*  
Fast foods and dining services exceed restaurants. That suggests the presence of a customer base that prefers drive-thru and carry-out restaurants. The data suggests also high presence of families. The density of the population for  $\text{km}^2$  is in mid-lower range (25-50 quantile). The average income is in a low-mid range (25 to 50 quantile). High concentration of bus stops suggests a population that need quick access to cheap transportation.

**2**  
*Very high density. Mid Income. Very high concentration of restaurants.*  
The cluster is characterized by very high residents' density. Income are in the mid-range. The concentration of restaurants is very high

**3**  
*Highest concentration of foreigners. Mid-high income. Mid density. Large offer for restaurants also foreign ones.*  
The cluster is characterized by the highest concentration of foreigners. Income are in the mid-high range. Residents density is in the mid-range. There is a large offer for restaurants also of foreign cuisines.



\* The optimal number of cluster was selected using the elbow method

# Neighbourhoods selection criteria

- **Success Factors for sustainable competitive advantage**

- Alignment between business idea and neighborhood environment
- Alignment of internal differentiating business capabilities with right market position
- Degree of rivalry in the neighborhood

- **Initial cluster selection**

- **Cluster 1 was excluded** as groups neighborhoods in which fast foods and dining services exceed restaurants, suggesting the presence of a customer base that prefers drive-thru and carry-out services

- **Neighborhoods scoring**

- The neighborhoods were scored considering the demographic factors, the pre-capita income, and the overall degree of rivalry

## Demographic Score

- People in that age range, have the highest pre-capita spending in restaurants
- One and two households have the highest pre-capita spending in restaurants
- Families with adult children have a moderate spending in restaurants

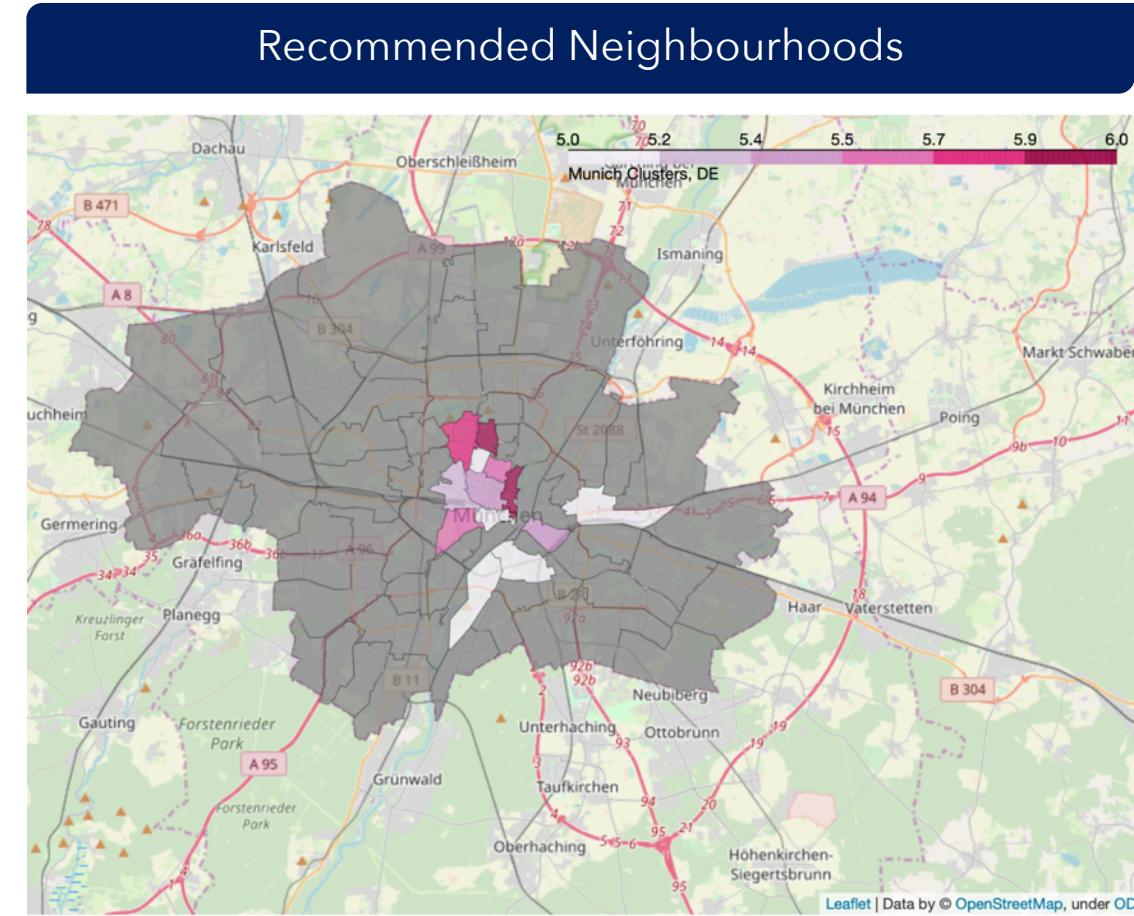
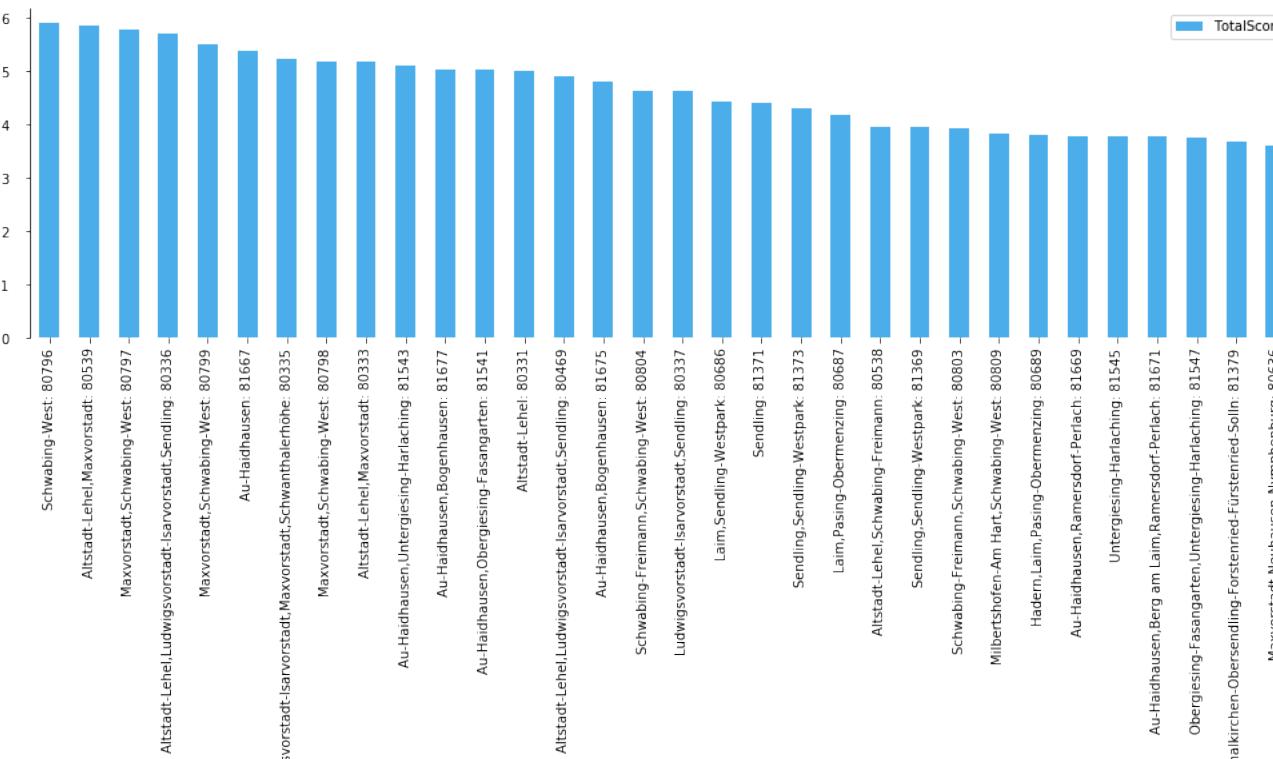
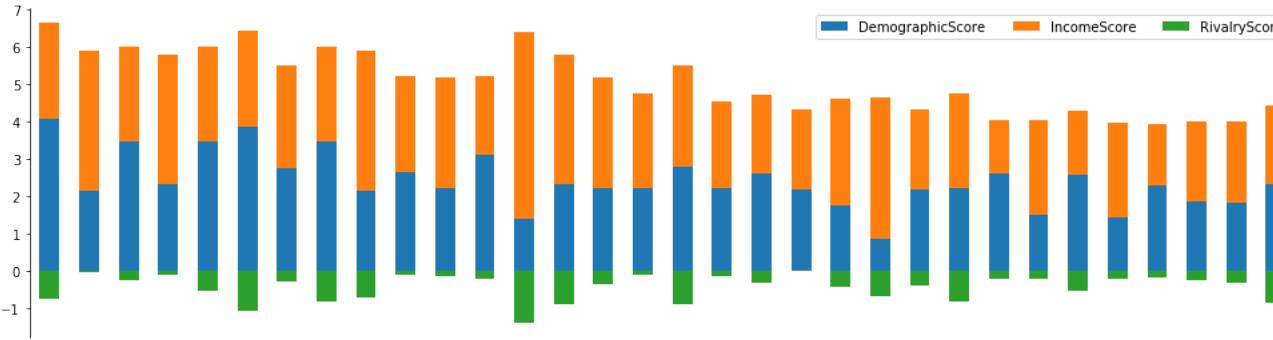
## Pre-Capita Income Score

- People with high income tend to spend more in restaurants than those with a lower one

## Rivalry Score

- The rivalry score gives a penalty based on the number of restaurants per adult, considering the top 4 culinary trends (Italian, German, Asian, Greek)

# Neighbourhoods scoring



# Conclusion and future work

---

- Provided a scoring model for the selection of a location for an Italian fine dining service
- The model developed for an Italian fine dining service, can be generalized, modifying the weighting scores according to the business marketing positioning strategy
- The data collection can be improved using multiple location data providers to broaden the set of venues.
- The scoring model can be refined including commercial real estate parameters, like rental costs, and demographic growth predictions

## AUTHOR NOTE

The case study presented in this document is solely an educational exercise. Any opinion, finding, conclusion, or recommendation is that of the author only and does not reflect by any means the view of any of the mentioned organizations, their employees or their administrations. The data disclosed in this document come from documentation, articles, and interviews publicly available or licensed to the author for this purpose. This document has been produced for the IBM Data Science Professional Certification Program only in order to evaluate and grade the knowledge of the author about the subjects of the 'Applied Data Science Capstone' learning module.

The author grants you a limited, revocable, nonexclusive, nontransferable license to view, store, bookmark, download and print the pages within this document and the related software solely for your personal, informational and noncommercial use, or as expressly authorized by the author in writing.

This case study and the related software are provided "as is", without warranty of any kind, express or implied, including but not limited to the warranties of merchantability, fitness for a particular purpose and noninfringement. in no event shall the authors or copyright holders be liable for any claim, damages or other liability, whether in an action of contract, tort or otherwise, arising from, out of or in connection with the software.