World's 50 Best Restaurants and Tourism

Carrer L.

10/10/2018

Introduction

Recently we have witnessing the growth of the so-called ?experience-based? tourism as a result of the shift in consumer?s value. Beside the common believe that food is one of the most important factor tourists take into account when it comes to travel destinations, there is also a wide literature confirming this belief. Although, food has always played been a critical factor in deciding the travel part of any holiday, the concept of traveling to a specific destination specifically for its restaurants scene is a recently new trend.

Moreover, in recent years we are experiencing a growth of the number of TV Programs, magazines and books publicizing the industry of food as a whole, ranging from the so-called ?street-food? to the high-end part of the business, which gather the gourmet restaurants and the niche of the ?Michelin-starred restaurants?. Therefore, such an exposure may have a positive effective in the decision-making of tourists.

The world?s 50 best restaurant list produced by UK Media Company William Reed Business Media reflect the heterogeneity of the culinary landscape while giving a snapshot of ?heute cuisine?. In fact while the list is mostly comprised from restaurants based in go-to destinations, there are also restaurants in off-beaten paths such as the Faviken in north-western Sweden, more than 600 Kilometers north from Stockholm. The guest-to-be may wonder why the chef has chosen this location, while others may wonder why they guests would decide to take such a journey only to visit a restaurant. This is also the aim of hospitality management studies: ?why tourist pick a travel destination? Do they factor the presence of such important culinary destinations??

To shed some light on these questions the following analysis has been conducted: does the number of arrivals for international tourism obtained from the World Bank website have a correlation with the presence of the world?s best 50 restaurants?

A limitation of this work is that data on arrivals are gathered at country level, therefore further studies could actually benefit from the usage of local tourism data. Moreover, while this analysis focus only on the existence of a correlation between the above-mentioned set of data, it would be interesting to assess in further studies the direction of the casual relationship between having various restaurants in the top 50 ranking and the tourism. Indeed, it is not clear whether tourists are attracted to destinations where one can find a top restaurant or the other way around, that is, restaurants are established where the market size is bigger due to tourists inflow.

Data Scraping

All the data for the project was scraped from the internet from three main sources: twitter.com, google-trends.com and repubblica.it. We are going to include the scraping code for all websites apart for twitter.com since it was a code written before the course and therefore does not pertain to this project. All websites are static and were scraped using python and different packages were used to perform the scraping. Repubblica.it is static and therefore was scraped using the requests package. Trends.google is a more complex website because of the restrictions imposed by Google. In order to speed up the scraping process, we used a third-part scraping package specifically designed for Google Trends: pytrends.request [1]. We are now going to speak about the different datasets more in detail.

All datasets cover a time span from 01/01/2015 to 01/08/2018. The twitter dataset consists in all the tweets from Italian populist party Te Northern League Matteo Salvini. For each tweet, the following characteristics

[1] http://ricerca.repubblica.it/ricerca/

are recorded: text, number of retweets, number of favorites, hashtags, mentions, geolocalization, id and link. The datasets consists in 15581 tweets in 1335 days for an tweets per day ratio of almost 12. Indeed, the analysis was motivated by the intense media presence of populist leaders.

The dataset from Repubblica.it[2] is meant to cover relevant news topics that have happened during the selected timeframe. Repubblica.it is the website of one of the majors italian newspapers, La Repubblica. The website is statics and has an archive page in which all past articles are collected. We exploit this webpage counting the number of results per serch during a 1-day timeframeThe data consists in daily counts of articles containing a specific word or word combination. Since the timespan considered is particularly large - 1335 days - we limit our search to the 13 words that were most relevant for our analysis.

Since we had a very limited capacity for what concerned the number of serches on Repubblica.it, we exploted the fact that Google trends can be scraped in a much more efficient fashion. The idea is the same: getting controls for relevant news or events that happened during the selected period. We recognize that Google Trends is not a perfect subsitute of a newspaper but it still captures most of the variation. The datasets consists in daily results form Google Trends for a long list of relevant words. The scale of each variable is artificial - it is normalized to have a peak at 100 - but it does not matter for our analysis since we are just interested in controlling for the variation in news and events.

The complete list of words is reported below. In bold are the words searched also on Repubblica.it, a small subset of the Google Trends list. The words are grouped in categories to control for different aspects of Salvini political agenda and/or news and events.

Populism Index

As we can observe in the following graph, the majority of the populist tweets fall into the "ostracizing the other category". This is expected since the main focus of the Northern League party is against immigration.

Descriptive statistics

Before proceding to the quantitative analysis we perform some exploratory analysis on the varaibles which we believe could be the most relevant driver for populism in Salvini's tweets. Because of the three populist categories, we choose three main news topics which could be most related to each of the categories. In particular, we analyze the following news variables: * news_migranti (# of articles with the word 'migrant' in it): supposedly captures events related to minorities and hence potential drivers of the 'ostracizing the others' channel. * news_tasse (# of articles with the word 'taxes' in it): as taxes are one of the most relevant expression of the power of the state on its citizens, it should capture postential drivers of the 'power to the people' channel. * news_unione_europea (# of articles with the composite word 'european union' in it): if we interpret the european union as an elitist organization, it should capture potential drivers of the 'against the elite' channel. * news_salvini(# of articles with the word 'salvini' in it): as a doublecheck.

From the graph above we observe some separation but the data is really noisy. In the next graph we check for separation in two dimensional spaces. We consider the spaces of each two-varaible combination where the list of variables is the same as above. From the graph above, we observe again little separation. There is no variable with a clear predictive power on its own. In the next section we explore the predictive power of news and events on the aggregate concentrating on all variables together. Since we have many variables with respect to our sample size (k/n ˜ 0.1), we use methods that involve some sort varaible selection: Lasso and Trees. Those methods allow us to explore both linear and nonlinear relations between our news data and the frequency/category of populist tweets.

[2] https://pypi.org/project/pytrends/

Results

For our second exercise, namely predicting the prevalence of daily populist tweets, we considered four outcome variables: for each of the three dimensions of populist ideology introduced earlier, we computed the percentage of daily tweets falling in that category. We also considered as an additional outcome the percentage of daily tweets which were categorized as populist, meaning those that matched either one of the three categories. The predictors considered in the analysis are given by the number of articles and lagged daily searches for the selected words/topics.
We start by presenting the Lasso path for each of our outcome variables, which were obtained using the glmnet package.

We then tuned the lambda parameter on a training dataset consisting of 70 per cent of our original dataset, using a 10-fold cross validation approach. The graphs below illustrate the evolution of the mean squared error over the grid of lambda values considered for the tuning of the parameter.

Cross-validation of lambda

As a parallel check we also tried tuning alpha and lambda for an elastic net model. We decided to present results from Lasso regressions (setting alpha to one) to focus on the sparsest set of predictors among the available covariates. If one wanted to instead focus on the predictive power of our model, then the tuning of both parameters should be considered. The following graph shows the evolution of the mean squared error for different values of alpha over the values of lambda.

Cross-validation alpha-lambda

We thus selected the lambda that minimized the mean squared error and estimated a Lasso model on the full dataset. The table below reports the sparse set of predictors selected for each of the four outcomes of interest.

Table 1: Correlation international tourist arrivals and number of restaurant

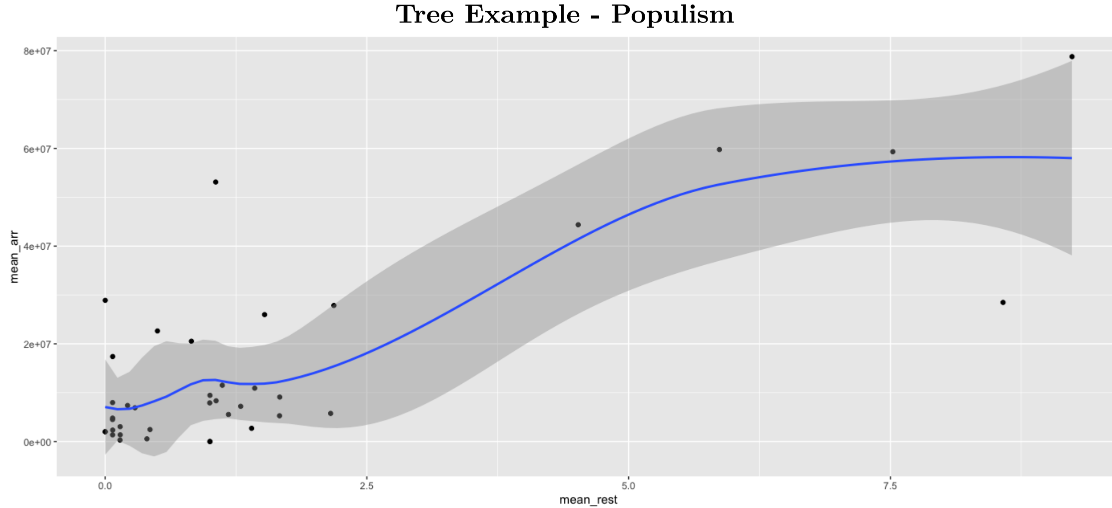
	Dependent variable:
	arrivals
num_restaurant	183,388.400
	(125,618.500)
Constant	1,033,445.000
	(962,115.200)
Year FE	Yes
Country FE	Yes
Observations	525
R^2	0.977
Adjusted R^2	0.975
Residual Std. Error	3,105,260.000 (df = 472)
F Statistic	391.177*** (df = 52; 472)
Note:	*p<0.1; **p<0.05; ***p<0.01

The We abstract from an interpretation of the signs of our parameters, since the sign of the Lasso estimate gives the sign of the correlation between the covariate and the residual. Nonetheless, there appear to be some interesting patterns:

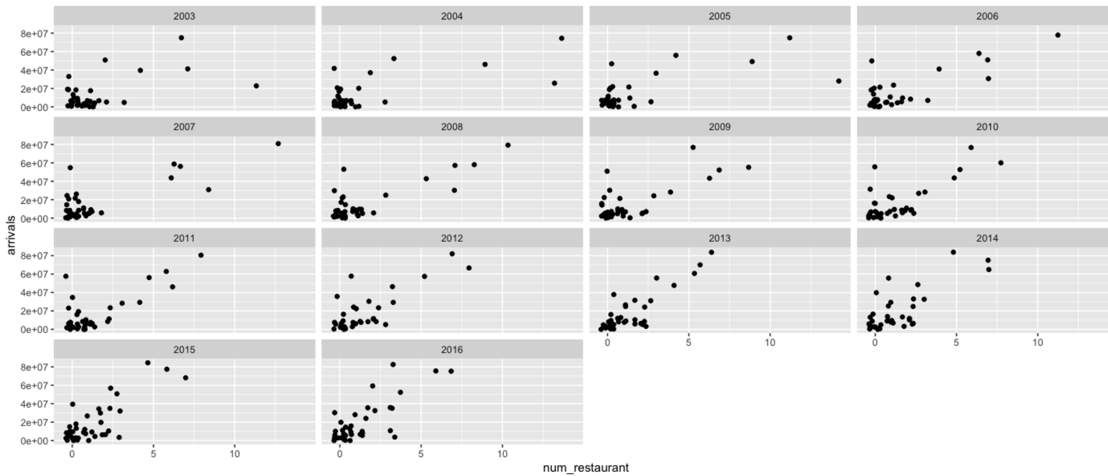
- among the strongest predictors positively associated with tweets belonging to the third category, "ostracizing the other", we find all the main words related to immigration (migrants, disembarking, illegal immigrants, Diciotti ship), in line with the main 'enemies within the people' in Salvini's tweets being immigrants and Roma communities;
- these same words show up also as predictors of populist tweets in general, which is consistent with a majority of such tweets pertaining to the third category and with a strategic exploitation of the news concerning the disembarking of migrants along the Italian coasts to portray the phenomenon of immigration as one of the main threats for the Italian people and economy;
- most of the predictors of tweets "against the ?lite" are of economic nature (deficit, austerity, "manovra finanziaria"), suggesting that attacks towards the elite, which are mainly addressed to the European Union, usually happen when economic reforms are discussed, perhaps in an attempt to frame the elites as responsible for the economic issues of the country;
- among the predictors of populist tweets we also find some economic policies often advocated by the LN (federalism, minimum income for citizens, removal of annuity payments for politicians) as well as words such as "cuts" and "waste";
- Another interesting result concerns the selection of "trend" variables related to opposition parties and political leaders. Tweets addressed "to the people" are correlated with peaks in searches of opposition leaders, such as Renzi, Di Maio, while populist tweets in general are correlated with searches for Lega or political actors close to LN along the political spectrum (e.g. Berlusconi, Meloni, Lega). This may suggest an intensifying of populist messages whenever the popularity of the party is falling;
- We also find correlation between the trending topic "electoral reimburses" and the incidence of populist tweets, suggesting that perhaps Salvini relied on the populist strategy more heavily when facing the scandal that involved his party.

We also tried to use trees and boosting to check whether the predictors selected with these other machine learning approaches were consistent with the results of the analysis performed with Lasso. The following graphs seem to suggest that this was the case.

Tree Example - Populism



Tree Example - Against the Elite



Tree Example - Power to the People

Tree Example - Ostracizing The Other

Summary of boosting splits - Populism

Summary of boosting splits - Against the Elite

Summary of boosting splits - Power to the People

Summary of boosting splits - Ostracizing The Other

Conclusion

In this project we have investigated correlations between the public display of populism and current news or events. We have used Matteo Salvini, the current leader of the italian party "The Northern League" as an example. From his dense twitter feed we have extrapolated a daily measure of populism and we have projected it against a wide variety of world events, italian news and general interest topics. Our result show that the data is very noisy but many interesting relationships arise. First of all, it seems that Salvini tweet feed is responsive to current news. For example, tweets against immigration re more likely to appear in days in which news report the landing of immigrants on italian soil. However, other non-intuitive correlations arise too. We cannot claim causation but temporal dimension supports causal speculations. We use as explanatory variables same-day news and previous-day trends which are supposedly pre-determined with respect to Salvini's tweet feed.