Which Retail Outlets Generate the Most Physical Interactions?

Avi Goldfarb and Catherine Tucker*

April 13, 2020

Abstract

This paper seeks to answer the simple question of what category of retail outlets generates the most physical interactions in the regular course of life. In this way, we aim to bring a marketing perspective to discussions about which businesses may be most risky from the standpoint of spreading contagious disease. We use detailed data from people's mobile devices prior to the implementation of social distancing measures in the United States. With this data, we examine a number of potential indicators of risk of contagion: The absolute number of visits and visitors, how many of the visits are generated by the same people, the median average distance traveled by the visitor to the retailer, and the number of customers from Canada and Mexico. We find that retailers with a single outlet tend to attract relatively few visitors, fewer one-off visitors, and have fewer international customers. For retailers that have multiple stores the patterns are non-linear. Retailers that have such a large number of stores that they are ubiquitous, tend to exhibit fewer visits and visitors and attract customers from a smaller distance. However, retailers that have a large enough footprint to be well known, but not large enough to be ubiquitous tend to attract a large number of visitors who make one-off visits, travel a long distance, and are disproportionately international.

*Avi Goldfarb is the Rotman Chair in Artificial Intelligence and Healthcare at the Rotman School of Management University of Toronto, and Research Associate at the NBER. Catherine Tucker is the Sloan Distinguished Professor of Marketing at MIT Sloan School of Management, Cambridge, MA and Research Associate at the NBER. All errors are our own

1 Introduction

COVID-19 has shut down large parts of the economy in order to save lives and increase long run economic output (Grinberg, 2007).[1] These non-pharmaceutical interventions have a long history in epidemiology (Heymann and Shindo, 2020), and there is recent evidence of effectiveness in the context of COVID-19 (Fang et al., 2020). At some point, it will be time to consider which parts of the economy to open first. Which businesses should remain open and which should shut? There is very little data to aid policymakers in these decisions (Aledort et al., 2007). That decision will partly depend on the economic benefit of opening those businesses with respect to employment, wages, and externalities on other industries. Perhaps more importantly, it will depend on whether there is a risk that opening those business will lead to a return to exponential growth of disease.

In this paper, we show that several key measures used in marketing analytics can also be used to establish which kind of businesses are likely to generate higher risk. To do this we highlight the evolution of several measures of retail traffic which have been developed in the digital era using location data from mobile phones. We employ February 2020 data from SafeGraph, which tracks over 40 million mobile devices in the US. This retail marketing data allows us to track the number of visits and unique visitors, median distance traveled to a retailer, the average time spent at a retail store, and number of foreign visitors across the month. By understanding the different types of customers that different retail businesses attract, we can provide suggestive guidance on the relative risk of opening businesses during an epidemic.

We focus on the question of whether small retailers or retail brands with multiple storefronts present the most risk. This is already a policy question in economies which are seeking to emerge from shutdown. For example, Austria has announced that it will allow small busi-

[1] For a survey of economists views see `http://www.igmchicago.org/surveys/policy-for-the-covid-19-crisis/7`

nesses to open before large businesses.[2] Given the liquidity constraints experienced by small retailers when faced with challenges (Dunne et al., 1988; Musso and Schiavo, 2008), this may seem to be an attractive type of business to open up early especially given recent survey evidence (Bartik et al., 2020), but in this paper we ask whether or not from a retail traffic perspective such a policy is justified.

We find that smaller retailers generally give rise to fewer physical interactions between people. Retailers with just a single store have fewer visitors and visits, and more repeat visitors and less one-off visits. Surprisingly, brands with a very large physical footprint (as measured by a large number of stores) also exhibit the same characteristics, and attract customers from the smallest relative distance. We suggest this is because brands with over 5000 outlets in the US are relatively ubiquitous, and therefore do not act as a magnet for shoppers to travel long distances to reach them. For example, Autozone, - a retailer that has over 5000 retail stores in the US, has a large footprint in order to minimize travel time for potential users, and does not act as a magnet for shoppers to make particularly long trips to its store. On the other hand, brands that have a mid-size footprint of between 10-1000 stores do attract more visitors and visits, and one-off visits, and customers who travel a longer distance to reach them. We speculate that this is because the brand is sufficiently strong to resonate in customers' minds, but the physical footprint is not large enough to make the store ubiquitous in the sense of being present in most American neighborhoods.

This data on retail traffic relates to the risk of contagion. For example, the Center for Disease Control (CDC) recommendations on postponing mass gatherings emphasize that larger gatherings "offer more opportunities for person-to-person contact and therefore post greater risk of COVID-19 transmission"[3]. The Public Health Agency of Canada adds that

[2] https://www.cnn.com/2020/04/11/health/european-countries-reopening-coronavirus-intl/index.html

[3] https://www.cdc.gov/coronavirus/2019-ncov/community/large-events/mass-gatherings-ready-for-covid-19.html

events that draw people from distant isolated communities generate greater risk.[4] More generally, it is well-established in the epidemiology literature that higher-density places where more people interact increase the likelihood of the spread of contagious disease (e.g. Fenichel (2013)).

This paper contributes to two strands of the management literature.

The first are economics and management papers that aim to guide policy during the time of global pandemic. These include papers have tried to use past history to guide policy (Barro et al., 2020), those that have studied the effectiveness of social distancing directives at halting the spread of the pandemic (Fang et al., 2020; Chiou and Tucker, 2020), and those that have tried to project out the likely business effects of the pandemic (Baker et al., 2020,?; Tucker and Yu, 2020). By contrast, this paper takes a marketing perspective and provides input into the specific forward-looking policy question of how to decide what a loosening of social distancing restrictions should look like.

The second are papers that have investigated the interaction between physical distance and retailing. Recently much of this literature has investigated the role of the internet and the 'death of distance' hypothesis (Kolko, 2000; Forman et al., 2009, 2018). Other work has studied how the digitization of supply chains has affected the role of distance in the supply of products to retail (Evans and Harrigan, 2005). Though much of this work has emphasized the idea that the internet has shifted the role of distance in retailing, we show that there is still substantial variation in the distance people travel to retailers, and that this seems to be systematically related to the nature of their physical footprint.

More broadly, this paper seeks to use mobile phone data to help guide a general policy framework about how to use mobile phone tracking data to estimate optimal public health responses. Phone data has already been widely used in the public health response to COVID-

[4] https://www.canada.ca/en/public-health/services/diseases/2019-novel-coronavirus-infection/health-professionals/mass-gatherings-risk-assesment.html

19. Phones have helped trace past contacts of infected individuals (Servick, 2020). Phone data has also been used to understand previous epidemics, such as the recent cholera epidemic in Haiti (Bengtsson et al., 2015). In marketing, papers have used mobile phone tracking data to try and optimize promotions and targeting (Ghose et al., 2019). Our analysis is specific to retail establishments, but the measures we analyze would be useful inputs into a more general model of how opening up specific workplaces would impact disease transmission.

Overall, this paper provides suggestive data and an empirical framework for helping decision-makers with respect to a specific and suddenly important policy problem: Which businesses can be allowed to open and which should remain shut in the face of COVID-19?

We recognize that this guidance would be one piece of a more complete assessment of risk. While we are not epidemiologists, we believe our perspective provides a useful input into their models. Interpreting our results to suggest that small businesses should be allowed to open first requires a number of additional assumptions, including that spillovers from still-closed larger businesses do not reverse our result, and that small businesses will be as good as larger businesses at public health measures such as masks for employees, hand sanitizer, and non-touch doors. Our results are not equilibrium, and do not take into account spillovers to other workplaces or the potential heterogeneity of prevention practices across businesses. Instead, our contribution is to demonstrate that a marketing perspective focused on customer behavior and data-driven marketing analytics can inform epidemiological models that simulate the impact of opening up different parts of the economy.

2 Data

We use data provided by Safegraph for the purposes of studying the spread of coronavirus for February 2020. This data is built on a panel of 44,546,450 million devices that collect anonymous location data. This represents about 10 percent of devices in the United States.[5]

[5] `https://www.safegraph.com/blog/what-about-bias-in-the-safegraph-dataset`

Each of the users of these devices has given permission for their location to be tracked by a variety of mobile apps. Safegraph matches the location of these devices to a variety of locations of branded physical retail locations within the US, and its main business is focused on providing data on retail traffic to firms and analysts.[6] This is advantageous for our purposes, as Safegraph's business model means it has already done substantive work to check the accuracy of the data.

There are several variables we focus on in particular. First, we examine the number of visits and the number of unique visitors to a store. These are measured by the number of times any mobile device visits a store location, and the number of times a unique device visits the store location. As we use aggregated measures of devices and their behavior, we do not have any insight about what the person holding the device was doing at that venue. They may have been shopping, they may have been browsing, they may have been meeting someone, or working at that store. Our measure does not distinguish between these purposes, but given that in a pandemic the largest concern is whether people congregate, rather than why they are congregating, we believe this does not pose too large a challenge for our core objective, though it does restrict the interpretation of the analysis.

The second variable is the median distance that month that a device traveled to go to that retail outlet. This is measured from the place that Safegraph has assigned as the home of the device. The home of the device is assigned to the place where that device spent the most time overnight. Evidently there is the potential for measurement error in the allocation of the home of the device if someone works overnight or stays overnight at another person's house regularly. However, given that we are interested predominantly in distance traveled to the store by the device, rather than distance traveled from home to the store, we believe that this type of error will not be consequential from our measurement.

[6]Safegraph also shared data with researchers who are working on measuring the effects of the spread of COVID-19.

Third, we examine time spent at the store. The effect of time spent at a store on contagion is ambiguous. More time spent at the store increases the likelihood of contagion at that store. At the same time, it means that those people are less likely to visit a large number of other stores. In other words, if non-branded stores have customers who spend more time, then it is unlikely that those customers visit a larger number of stores. We display the results, but a full assessment of the implications of time spent requires additional data and a more complete model.

Fourth, we examine devices that appear to be based in Canada and Mexico but travel to the US regularly. These are a small fraction of the data, and when we analyze these visits by foreign devices we focus on the states bordering each country in the US, where these visits tend to happen. We include this as an incomplete measure of the degree to which foreign nationals tend to visit stores within the US.

We use data from February 2020. We choose this period as this was before people in the US shifted behavior towards self-isolation. On February 29th, the US experienced its first COVID-19 death. Our intention therefore with this selection of data is that it provides the last glimpse of normal human traffic patterns prior to the shifts in behavior due to intentional social distancing. We look at the entire month of February, and use data on over 1 million commercial retail locations in the US, for which Safegraph had allocated a NAICS code. To look at retailers we focus on all locations which had the 2-digit NAICS code of 44-45 which spans the category of 'Retail Trade'.

This dataset is focused on the US. Two avenues exist whereby the data is potentially not representative of behavior in February 2020. First, it does not represent behavior of people who do not have smartphones. Another source of bias is that the people who opt in to allow their location to be tracked may not be representative of the population. Goldfarb and Tucker (2012) show that willingness to divulge personal information decreases with age. Athey et al. (2017) suggests that to some extent the decision to divulge information is highly contextual and can be easily shifted through small incentives and changes in interfaces, implying that at least potentially some randomness in the decision to divulge may exist. Safegraph suggest that their data does align with Census data.[7]

Table 1 provides summary statistics for the Safegraph data. It suggests that the average distance traveled is 19.6 km to the retail store in our data, but the median is 8.6 km. On average people spent 30 minutes in a store, with a median of 20 minutes. The average store had 176 separate visitors and 283 visits during the month from devices in the Safegraph data.[8]. The Visitor to Visit Ratio measures the extent to which the same visitor made multiple visits. A value of 1 indicates that no visitors made repeat visits. A lower value indicates that visitors tended to make repeat visits throughout the month.

Table 1 also provides some insights into the nature of the retail stores in our data. The average number of outlets is 1221.4, but this is highly skewed. The median number of oulets is 1, with 52% of the data having a single outlet. To deal with the skewed nature of this variable in our analysis, we created a series of binary indicators which indicate what bucket in terms of number of outlets that particular retail chain belongs to. Most of these buckets consist of around 10% of our observed data, with the exception of the indicator for 10-99 outlets which is 6% of our observed data.

11% of the outlets are gas stations. 14% of outlets are devoted to food. 3% are drug stores. We separate out gas stations, pharmacies and food retailers from the rest of our retailers in our dataset in some of our analysis, as these are retailers that are usually marked as 'essential businesses' whose opening is not in question.

The specific stores that tended to exhibit the greatest distance traveled were not particular surprising. For example, 'Pilot Flying J,' a chain of truck stops had a median distance

[7] `https://colab.research.google.com/drive/1u15afRytJMsizySFqA2EPlXSh3KTmNTQ#offline=true&sandboxMode=true`

[8] Given that Safegraph has about 10 percent of devices in the US, one way to get back-of-the-envelope total visits would be to multiply by ten

Table 1: Summary Statistics

	Mean	Median	Std Dev	Min	Max
Distance (km)	19.6	8.57	124.9	0.0020	6561.0
Average Time Spent	31.2	20	41.3	4	1429.5
Visitor Count	176.0	74	371.3	5	39996
Visit Count	282.8	125	637.9	5	72219
Visitor to Visit Ratio	0.66	0.67	0.20	0.016	1
Canadian Visitors	0.0033	0	0.057	0	1
Mexican Vistors	0.0061	0	0.078	0	1
> 5000 Outlets	0.098	0	0.30	0	1
1000-4999 Outlets	0.11	0	0.31	0	1
100-999 Outlets	0.12	0	0.32	0	1
10-99 Outlets	0.055	0	0.23	0	1
2-9 Outlets	0.100	0	0.30	0	1
Single outlet	0.52	1	0.50	0	1
Outlets	1221.4	1	2902.4	1	15718
Food Store	0.14	0	0.35	0	1
Gas Station	0.11	0	0.32	0	1
Drug Store	0.033	0	0.18	0	1

traveled to them of 379km. ABC Stores - a chain of stores that targets tourists in Hawaii - had an average median distance of 2329km. The stores that exhibited the greatest number of visitors were also not surprising - Walmart accounts for 73% of all stores that had over 10,000 visits.

Table 2 breaks down these summary statistics by 3-digit NAICS codes for the different types of stores in our data. It is clear that there are large differences by type of store. Clothing stores attract people from the furthest distance. General stores (which encompass brands such as Walmart and Target) attract the most visits and visitors. Home stores have the most repeat customers, while clothing stores have the fewest repeat visitors. Stores with only a single outlet are most concentrated in the leisure category, which spans music, books, sports, and hobbies. In the appendix Figures A1, A2 and A3, we provide further breakdowns of Table 2, for single-outlet, mid-tier, and large footprint brands. There are no brands of stores in our data that have over 5000 outlets in the Furniture, Misc, Gas, or

Table 2: Summary Statistics By Different Store Types

	Auto	Clothing	Electronics	Food	Furniture	Gas	General	Health	Home	Leisure	Misc
Visit Count	147.4	165.2	160.8	334.8	137.0	330.7	967.4	225.6	196.8	279.5	249.6
Visitor Count	81.44	116.4	95.59	197.9	86.94	221.5	591.2	148.5	114.7	177.5	157.4
Visitor to Visit Ratio	0.604	0.768	0.632	0.654	0.660	0.689	0.664	0.706	0.588	0.638	0.645
Distance (km)	17.36	35.26	14.54	17.93	14.62	25.37	13.34	11.92	13.24	27.80	17.98
Average Time Spent	34.15	31.43	41.94	25.42	36.67	14.87	22.32	30.77	40.07	40.93	36.14
Single outlet	0.544	0.502	0.686	0.528	0.588	0.229	0.154	0.519	0.627	0.740	0.675
Total	148620	89986	25842	155189	51069	124300	67277	98730	87394	98047	155089
Observations	1101543										

Leisure categories, so we also checked the robustness of our results to their exclusion.

3 Empirical Analysis

Figure 1 and 2 look at the number of visits and vistors by number of outlets. They show similar patterns, in that brands that only have a single storefront receive fewer visits and visitors. The number of visits and visitors peaks at stores that have between 1000-5000 outlets.

Because there are differences across location and store-type that this graphical evidence does not take into account, we shift to regressions. Our empirical specification is straightforward. It presents the descriptive statistics from the figures above, but in regression form in order to control for zip code and type of store. For store i of type k in zipcode z, we model visitation pattern as:

$$VisitationPattern = \beta NumberofOutlets + \alpha_k + \gamma_z + \epsilon_z$$

Our focus is on how the stores's physical footprint, or number of individual outlets across

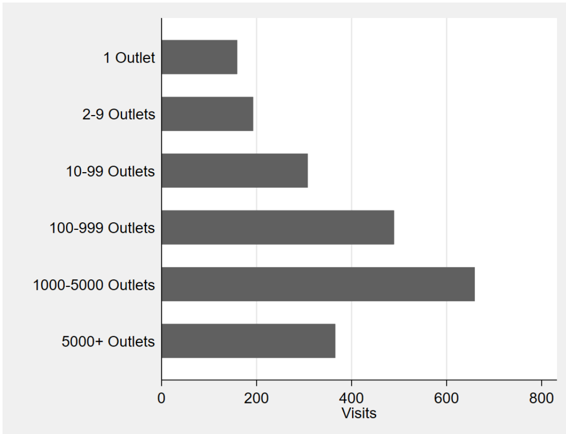

Figure 1: Number of Visits

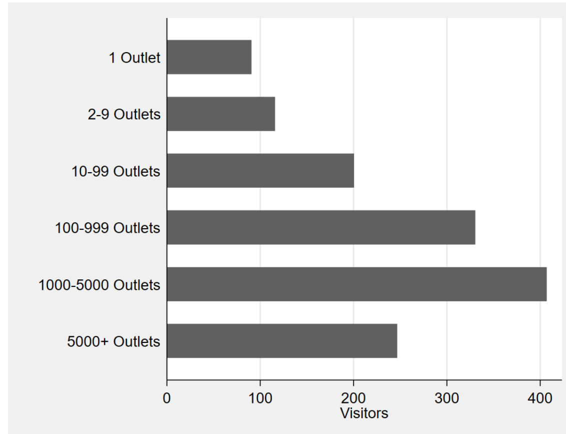

Figure 2: Number of Visitors

the US, affects visitation patterns. α_k is a vector of fixed effects for the different six-digit NAICS codes of stores in our data. γ_z is a vector of fixed effects for each zipcode.

Table 3 examines how the extent of a brand's physical store footprint affects visitation patterns. More visits and more visitors likely mean more risk of contagion. Columns (1)-(3) look at the entire data set, and Columns (4)-(6) excludes Gas Stations, Food Stores, and Drug Stores. The motivation for this exclusion is that most stay-at-home orders exclude pharmacies, gas stations, and food stores from having to shut because they are essential services. Therefore, given our focus is on helping policy makers understand which stores to open first, it makes sense to look at stores which are shut as a result of a variety of state directives.

There is a reasonably consistent pattern that stores with just one outlet (the excluded group) attracted fewer visitors and visits, while stores with more than 1000 outlets but fewer than 5000 outlets tend to attract the most visitors and the most visits. In addition, stores with the largest number of outlets attract relatively few visitors. These large stores also exhibit fewer visits than single storefronts when we control for location and store-type. In order to check that this pattern was not driven by differences in number of outlets across categories, we confirmed robustness to excluding categories such as electronics and clothing

Table 3: Stand-Alone Retailers and Large Footprint Retailers Attract the Smallest Number of Visitors

	All	Non-Essential Businesses				
	(1)	(2)	(3)	(4)	(5)	(6)
	Visitor Count	Visit Count	Visitor to Visit Ratio	Visitor Count	Visit Count	Visitor to Visit Ratio
> 5000 Outlets	22.14***	-32.35***	0.0845***	-230.9***	-504.2***	0.109***
	(1.752)	(3.129)	(0.000755)	(4.368)	(8.074)	(0.00116)
1000-4999 Outlets	262.4***	426.2***	0.0836***	339.9***	567.1***	0.0962***
	(2.161)	(3.727)	(0.000719)	(2.939)	(5.149)	(0.000878)
100-999 Outlets	221.5***	306.2***	0.104***	147.8***	170.6***	0.130***
	(1.918)	(2.965)	(0.000803)	(2.154)	(3.134)	(0.000906)
10-99 Outlets	96.18***	130.3***	0.0718***	71.10***	82.74***	0.0832***
	(1.786)	(2.781)	(0.000943)	(2.189)	(3.312)	(0.00123)
2-9 Outlets	20.60***	26.47***	0.0262***	21.37***	26.88***	0.0298***
	(0.728)	(1.255)	(0.000674)	(0.878)	(1.508)	(0.000783)
Zipcode Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
NAIC Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1102358	1102358	1102358	783889	783889	783889
R-Squared	0.225	0.195	0.193	0.259	0.235	0.217

Dependent variable is the number of visitors to the store in Columns (1) and (4). Dependent variable is the number of visits to the store in Columns (2) and (5). Dependent variable is the ratio of visitors to visits in Columns (3) and (6) * p <0.10, ** p <0.05, *** p <0.01. Data for February 2020. Robust standard errors clustered at the zipcode level. * p <0.10, ** p <0.05, *** p <0.01.

where chains did not have more than 1000 outlets.

Figure 3 shows that the stores with the largest number of outlets tend to have customers who travel the least distance and who spend the least time. Stores with 10-99 outlets have a relatively large distance traveled.

Table 4 presents these descriptive statistics on distance traveled and time spent in regression form. In Columns (1) and (3) we present results for the full sample. In Columns (2) and (4) we present results for non-essential businesses only.

The results suggest that similar to Figure 3, there is a non-linear pattern in terms of how far people travel to visit stores. In general, if a brand is large enough to be relatively ubiquitous and has hundreds of stores, then people are less likely to travel far to visit. Presumably by the time a brand has 5000 outlets it is likely trying to minimize travel time for its potential customers. By contrast brands where there are fewer than 100 outlets (but more than a single outlet) are more likely to attract people to drive a greater distance.

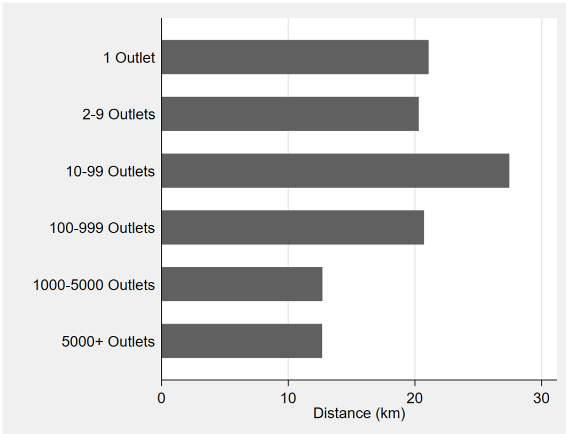

Figure 3: Distance Traveled to Stores

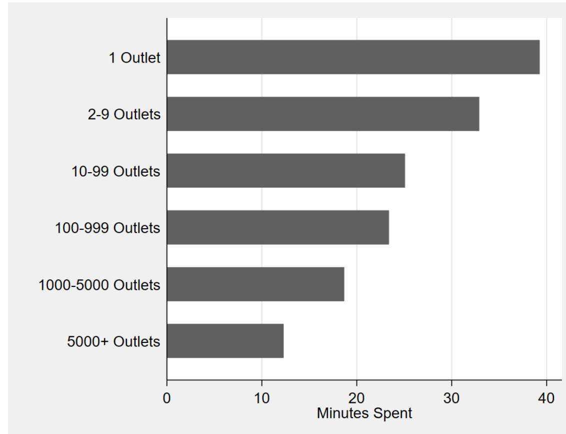

Figure 4: Time Spent in Stores

Table A4 in the appendix reports a robustness check where we exclude stores in Hawaii and Alaska since they might be expected (given the distances tourists travel to them) to have particularly long median distances traveled. The general pattern of results remains similar.

The results for time spent, by contrast, suggest a linear pattern in how a brand's physical footprint affects how long people stay there. People are most likely to spend longest at a retail store with a single storefront. By contrast, as a brand's retail store footprint becomes ubiquitous, people are less likely to stay long at the outlet. Therefore, it is unlikely that visitors to these single outlet stores visit a larger number of stores.

Table 5 looks at how foreign visitors to a store vary with the store's footprint. In general, Canadian visitors are more likely to visit stores with a midsize footprint and less likely to visit stores that are either more ubiquitous or a single storefront. One explanation of this is simply that these visits reflect Canadians crossing the border in attempt to shop at stores with lower prices than those found in Canada, and that price-orientated stores that do not exist in Canada tend to have these mid-price footprints. There are somewhat similar patterns for Mexican visitors, but many of our estimates are non-significant.

Table 4: Mid-Tier Brands Attract People from the Greatest Distance

	All	Non-Essential Businesses	All	Non-Essential Businesses
	(1)	(2)	(3)	(4)
	Distance (km)	Distance (km)	Average Time Spent	Average Time Spent
> 5000 Outlets	-6.374***	-8.332***	-19.57***	-24.14***
	(0.389)	(0.534)	(0.132)	(0.209)
1000-4999 Outlets	-4.971***	-3.999***	-17.11***	-18.35***
	(0.328)	(0.388)	(0.123)	(0.156)
100-999 Outlets	2.238***	-1.454***	-15.00***	-15.03***
	(0.411)	(0.410)	(0.120)	(0.148)
10-99 Outlets	2.668**	5.129***	-11.31***	-10.80***
	(0.836)	(1.162)	(0.159)	(0.216)
2-9 Outlets	0.638	1.319**	-4.881***	-5.207***
	(0.376)	(0.498)	(0.143)	(0.165)
Zipcode Fixed Effects	Yes	Yes	Yes	Yes
NAIC Fixed Effects	Yes	Yes	Yes	Yes
Observations	1098838	780960	1102358	783889
R-Squared	0.413	0.455	0.120	0.106

Dependent variable is median distance traveled in kilometers to the store in Columns (1)-(2). Dependent variable is the median length of time spent in the store in Columns (3)-(4). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Data for February 2020. Robust standard errors clustered at the zipcode level.

Table 5: Stand-Alone Stores And Large Footprint Brands Appear to Attract the Fewest Canadian Visitors

	All (1) Canadian Visitors	Non-Essential Businesses (2) Canadian Visitors	All (3) Mexican Vistors	Non-Essential Businesses (4) Mexican Vistors
> 5000 Outlets	-0.00328***	-0.00237***	-0.000517*	-0.000786
	(0.000383)	(0.000511)	(0.000206)	(0.000406)
1000-4999 Outlets	-0.00191***	0.00142***	-0.000133	-0.0000793
	(0.000375)	(0.000409)	(0.000138)	(0.000187)
100-999 Outlets	0.00678***	0.000224	0.00104**	0.000676
	(0.000709)	(0.000287)	(0.000352)	(0.000425)
10-99 Outlets	0.000989**	0.000598	0.000418	0.000687
	(0.000361)	(0.000445)	(0.000307)	(0.000487)
2-9 Outlets	0.000131	0.000143	0.0000666	0.000112
	(0.000216)	(0.000214)	(0.000186)	(0.000247)
Zipcode Fixed Effects	Yes	Yes	Yes	Yes
NAIC Fixed Effects	Yes	Yes	Yes	Yes
Observations	257964	187958	239285	167130
R-Squared	0.0626	0.0504	0.0878	0.107

Dependent variable is whether the store received visitors from Canada. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Sample restricted to states that border Canada. Data for February 2020. Robust standard errors clustered at the zipcode level.

4 Conclusion

This paper uses metrics of retail foot-traffic which are commonly used in marketing analytics, to provide policymakers with suggestive evidence about what kind of businesses can be most safely opened up when the economy emerges from directives that closed down all but essential businesses. We demonstrate that such metrics may be a useful input into policy models which evaluate which businesses are safest to allow to reopen in the wake of a global pandemic.

In this paper we present evidence that is generally supportive of a policy which enables small businesses, particularly stores with a single retail storefront, to open first. These stores attract fewer visits and fewer visitors, and more repeat customers. However, their customers do travel a longer distance and spend longer in the store relative to stores whose brands are relatively ubiquitous.

There are limitations to our paper. First, we look simply at mobile tracking data. We do not have any information about what types of activities (working, browsing, or purchasing) people took in these stores. Second, we do not try and make a relative assessment of which of the factors we consider - distance traveled, time spent, number of visits, number of visitors and visitor to visit ratio - is most important to prioritize when making decisions about which stores to allow to open first. Third, we have very limited behavior on the pattern of people who have visited foreign countries, and the data we have is unlikely to be representative of the behavior of such people in general, as it appears we are mainly capturing the actions of people who cross the border to save money by shopping in the US. Fourth, data from February may not fully represent traffic patterns during the time period that policy-makers are considering re-opening businesses. Notwithstanding these limitations, we believe that this paper is a useful step for trying to set up a framework to evaluate what kind of businesses should open first, providing a marketing perspective to a key input into models of non-pharmaceutical interventions.

References

Aledort, J. E., N. Lurie, J. Wasserman, and S. A. Bozzette (2007). Non-pharmaceutical public health interventions for pandemic influenza: an evaluation of the evidence base. *BMC public health 7*(1), 208.

Athey, S., C. Catalini, and C. Tucker (2017, June). The digital privacy paradox: Small money, small costs, small talk. Working Paper 23488, National Bureau of Economic Research.

Baker, S. R., N. Bloom, S. J. Davis, K. J. Kost, M. C. Sammon, and T. Viratyosin (2020, April). The Unprecedented Stock Market Impact of COVID-19. Working Paper 26945, National Bureau of Economic Research.

Baker, S. R., R. Farrokhnia, S. Meyer, M. Pagel, and C. Yannelis (2020, April). How Does Household Spending Respond to an Epidemic? Consumption During the 2020 COVID-19 Pandemic. Working Paper 26949, National Bureau of Economic Research.

Barro, R. J., J. F. Ursúa, and J. Weng (2020, March). The coronavirus and the great influenza pandemic: Lessons from the "spanish flu" for the coronavirus's potential effects on mortality and economic activity. Working Paper 26866, National Bureau of Economic Research.

Bartik, A. W., M. Bertrand, Z. B. Cullen, E. L. Glaeser, M. Luca, and C. T. Stanton (2020, April). How Are Small Businesses Adjusting to COVID-19? Early Evidence from a Survey. Working Paper 26989, National Bureau of Economic Research.

Bengtsson, L., J. Gaudart, X. Lu, S. Moore, E. Wetter, K. Sallah, S. Rebaudet, and R. Piarroux (2015). Using mobile phone data to predict the spatial spread of cholera. *Scientific reports 5*, 8923.

Chiou, L. and C. E. Tucker (2020). Social distancing, internet access and inequality. *Mimeo, MIT*.

Dunne, T., M. J. Roberts, and L. Samuelson (1988). Patterns of firm entry and exit in us manufacturing industries. *The RAND journal of Economics*, 495–515.

Evans, C. L. and J. Harrigan (2005). Distance, time, and specialization: Lean retailing in general equilibrium. *American Economic Review 95*(1), 292–313.

Fang, H., L. Wang, and Y. Yang (2020, March). Human mobility restrictions and the spread of the novel coronavirus (2019-ncov) in china. Working Paper 26906, National Bureau of Economic Research.

Fenichel, E. P. (2013). Economic considerations for social distancing and behavioral based policies during an epidemic. *Journal of health economics 32*(2), 440–451.

Forman, C., A. Ghose, and A. Goldfarb (2009). Competition between local and electronic markets: How the benefit of buying online depends on where you live. *Management science 55*(1), 47–57.

Forman, C., A. Goldfarb, and S. Greenstein (2018). How geography shapes-and is shaped by-the internet. *The New Oxford Handbook of Economic Geography 269*.

Ghose, A., B. Li, and S. Liu (2019). Mobile targeting using customer trajectory patterns. *Management Science 65*(11), 5027–5049.

Goldfarb, A. and C. Tucker (2012). Shifts in privacy concerns. *American Economic Review: Papers and Proceedings 102*(3), 349–53.

Grinberg, M. (2007). Learning from the 1918 pandemic. *Risk Management 54*(8), 10.

Heymann, D. L. and N. Shindo (2020). COVID-19: what is next for public health? *The Lancet 395*(10224), 542–545.

Kolko, J. (2000). The death of cities? The death of distance? Evidence from the geography of commercial Internet usage. *The internet upheaval: Raising questions, seeking answers in communications policy*, 73–98.

Musso, P. and S. Schiavo (2008). The impact of financial constraints on firm survival and growth. *Journal of Evolutionary Economics 18*(2), 135–149.

Servick, K. (2020, March 22). Cellphone tracking could help stem the spread of coronavirus. Is privacy the price? *Science*.

Tucker, C. E. and S. Yu (2020). The early effects of coronavirus-related social distancing restrictions on brands. *Available at SSRN 3566612*.

Table A1: Summary Statistics By Different Store Types (Single Stores)

	Auto	Clothing	Electronics	Food	Furniture	Gas	General	Health	Home	Leisure	Misc
Visit Count	107.3	136.9	133.3	162.7	108.4	173.8	253.8	160.5	99.54	214.6	204.4
Visitor Count	55.72	84.18	67.74	92.48	58.28	103.8	149.0	96.35	49.04	119.7	120.8
Visitor to Visit Ratio	0.588	0.700	0.586	0.638	0.603	0.645	0.609	0.674	0.552	0.606	0.622
Distance (km)	19.82	32.46	14.56	21.08	15.25	23.07	31.16	12.29	13.53	30.49	20.05
Average Time Spent	38.41	37.33	47.83	32.24	42.43	24.51	42.49	38.83	47.99	44.92	39.30
Total	148620	89986	25842	155189	51069	124300	67277	98730	87394	98047	155089
Observations	577896										

Table A2: Summary Statistics By Different Store Types (2-1000 outlets)

	Auto	Clothing	Electronics	Food	Furniture	Gas	General	Health	Home	Leisure	Misc
Visit Count	199.2	193.8	220.9	528.5	181.2	493.0	1038.7	194.1	177.2	448.4	306.1
Visitor Count	109.3	148.9	156.4	313.0	128.8	331.1	719.1	127.7	106.2	326.2	204.6
Visitor to Visit Ratio	0.601	0.836	0.732	0.665	0.729	0.690	0.735	0.731	0.619	0.716	0.683
Distance (km)	17.57	38.08	14.51	16.28	14.23	44.04	17.00	15.98	14.57	22.44	14.97
Average Time Spent	34.42	25.48	29.08	20.23	29.43	15.26	24.91	29.94	32.80	31.65	31.19
2-9 Outlets	0.374	0.200	0.308	0.346	0.332	0.298	0.140	0.523	0.529	0.461	0.551
10-99 Outlets	0.147	0.205	0.122	0.284	0.212	0.272	0.148	0.192	0.183	0.140	0.196
100-999 Outlets	0.479	0.594	0.570	0.370	0.456	0.430	0.712	0.284	0.287	0.399	0.253
Total	148620	89986	25842	155189	51069	124300	67277	98730	87394	98047	155089
Observations	297051										

Table A3: Summary Statistics By Different Store Types (1000+ outlets)

	Auto	Food	Furniture	Gas	General	Health	Home	Leisure	Misc
Visit Count	189.0	524.3	151.0	340.1	1113.9	373.8	530.2	536.2	469.3
Visitor Count	116.8	322.7	119.9	232.4	658.9	263.8	335.4	413.8	330.8
Distance (km)	9.428	9.875	9.678	20.28	8.175	8.107	11.06	9.936	9.487
Visitor to Visit Ratio	0.659	0.691	0.830	0.705	0.657	0.746	0.675	0.796	0.726
Average Time Spent	20.45	11.86	20.63	10.97	16.91	16.05	21.13	20.12	24.17
1000-4999 Outlets	0.584	0.638	1	0.410	0.312	0.396	1	1	1
$>$ 5000 Outlets	0.416	0.362	0	0.590	0.688	0.604	0	0	0
Total	148620	155189	51069	124300	67277	98730	87394	98047	155089
Observations	226596								

Table A4: Mid-Tier Brands Attract People from the Greatest Distance (Excluding Hawaii and Alaska)

	All	Non-Essential Businesses	All	Non-Essential Businesses
	(1)	(2)	(3)	(4)
	Distance (km)	Distance (km)	Average Time Spent	Average Time Spent
> 5000 Outlets	-5.993***	-7.210***	-19.56***	-24.14***
	(0.225)	(0.240)	(0.133)	(0.209)
1000-4999 Outlets	-4.295***	-3.083***	-17.11***	-18.37***
	(0.186)	(0.173)	(0.124)	(0.157)
100-999 Outlets	2.945***	-0.776**	-15.01***	-15.05***
	(0.315)	(0.255)	(0.120)	(0.148)
10-99 Outlets	1.659***	3.220***	-11.31***	-10.80***
	(0.407)	(0.548)	(0.160)	(0.217)
2-9 Outlets	0.976***	1.602***	-4.863***	-5.196***
	(0.288)	(0.367)	(0.143)	(0.165)
Zipcode Fixed Effects	Yes	Yes	Yes	Yes
NAIC Fixed Effects	Yes	Yes	Yes	Yes
Observations	1091956	775734	1095447	778639
R-Squared	0.325	0.366	0.121	0.106

Dependent variable is median distance traveled in kilometers to the store in Columns (1)-(2). Dependent variable is the median length of time spent in the store in Columns (3)-(4). * $p<0.10$, ** $p<0.05$, *** $p<0.01$. Data for February 2020. Robust standard errors clustered at the zipcode level.