MIT 11.521/524 Class Project Report, Spring 2012

Title:

Transit-Oriented Development: A Destination-Based Analysis

Abstract:

Historically, evaluations of transit-oriented developments have focused on quantifying nearby residential population and that population's transit use. This evaluation seeks to analyze TOD from a new perspective by examining the potential of destinations to attract transit riders to a particular stop. The study focuses on 18 subway stations identified by the Metropolitan Area Planning Commission as Neighborhood Subways. Using 2009-2010 data from the MBTA, ridership profiles for the 18 stations were created and then 6 were identified for more thorough case study analyses. Assessment of the urban form within 800 meters of the stations was conducted using MassGIS. This assessment identified the infrastructure assets and challenges at each stop, such as wide sidewalks or highway barriers. Lastly, business data from InfoGroup was combined with data from the social media platform Foursqaure to provide an analysis of business type, sales, and popularity of various destinations within 800 meters of each station. The analysis concludes with detailed descriptive profiles of the 6 case study stations. The robust profiles developed from the different data sets underscores the importance of incorporating new data sources and understanding the destination impact on TODs.

Participant List:

Alexis Howland
Alstan Jakubiec
Jonathan Ragan-Kelley
Elizabeth Resor
RJ Williams

Introduction

Historically, evaluations of transit-oriented developments have focused on quantifying nearby residential population and that population's transit use. This evaluation, however, seeks to analyze TOD from a new perspective by examining the potential of destinations to attract transit riders to a particular stop. We believe that the current designations of Boston MBTA stations are not nuanced enough to completely demonstrate the full spectrum of stations within the system. This evaluation reveals that as transit-oriented development moves forward both the origin and the destination side of stations should be considered before a station is labeled a success or failure. Additionally, this in-depth inspection will lead to more specific paths for improvements of stations that might have been previously missed if planners relied on a one-size-fits-all approach to redevelopment. In order to respond to our hypothesis of destination-based development, the study examined the stations on the axes of ridership, urban form and economic activity.

Methodology

Using 2009-2010 data from the MBTA, ridership information contains more than the percent of the population captured (although that was a metric that was considered). Also included are issues such as the spread of the data (i.e., was there a peak at morning commute or was it evenly spread throughout the day?), percent of weekday ridership captured on the weekend and other breakdowns such as the age of rider and what type of ticket they used. Urban form was included to test the hypothesis that stations' pedestrian accommodations are a line of differentiation, and that 'good' urban form will create more attraction to riders given a similar mix of destination activities. The metrics for urban form within an 800 meter radius of the station was conducted using MassGIS. The components for urban form analysis included the ratio of sidewalk area to road area, the number of intersections within each station boundary (to reflect the Boston tradition of Squares and Plazas), and also examined whether roads are seen as pathways to destinations within the boundary or if they acted as barriers and divided a station from its attraction (this classification was decided by the width of the road, its speed limit and its classification within the Massachusetts EOT Roads data). Lastly, we looked at economic activity which we defined as the number of Foursquare CheckIns, which showed the business type and popularity of a site, and InfoGroup sales data which we classified by type of business in an attempt to normalize for different average transaction prices.

To break the network of MBTA stations into a manageable subset, we decided to work with what the Metropolitan Area Planning Commission (MAPC) designates "Neighborhood Subways," which are defined as "subway and trolley station areas in predominantly residential, moderate-density neighborhoods throughout the Inner Core." To further hone in on the stations (which started with 63 stations in the classification) in the study, the Green Line stations were removed from the study for two reasons: 1. The ridership data received only had daily totals from each Green Line station, as opposed to the hourly breakdown that the other lines had; and 2. The Green Line stations function in a manner

similar to trolley stations so that they are often very close together and the numbers received were often small in comparison to the other stations. The eighteen stations are identified in the table below.

18 Neighborhood Stations	Subway Line
Andrew Square	Red
Ashmont	Red
Beachmont	Blue
Central Square	Red
Community College	Orange
Davis	Red
Fields Corner	Red
Green Street	Orange
Jackson Square	Orange
Orient Heights	Blue
Porter Square	Red
Revere Beach	Blue
Roxbury Crossing	Orange
Savin Hill	Red
Shawmut	Red
Stony Brook	Orange
Wollaston	Red
Wood Island	Blue

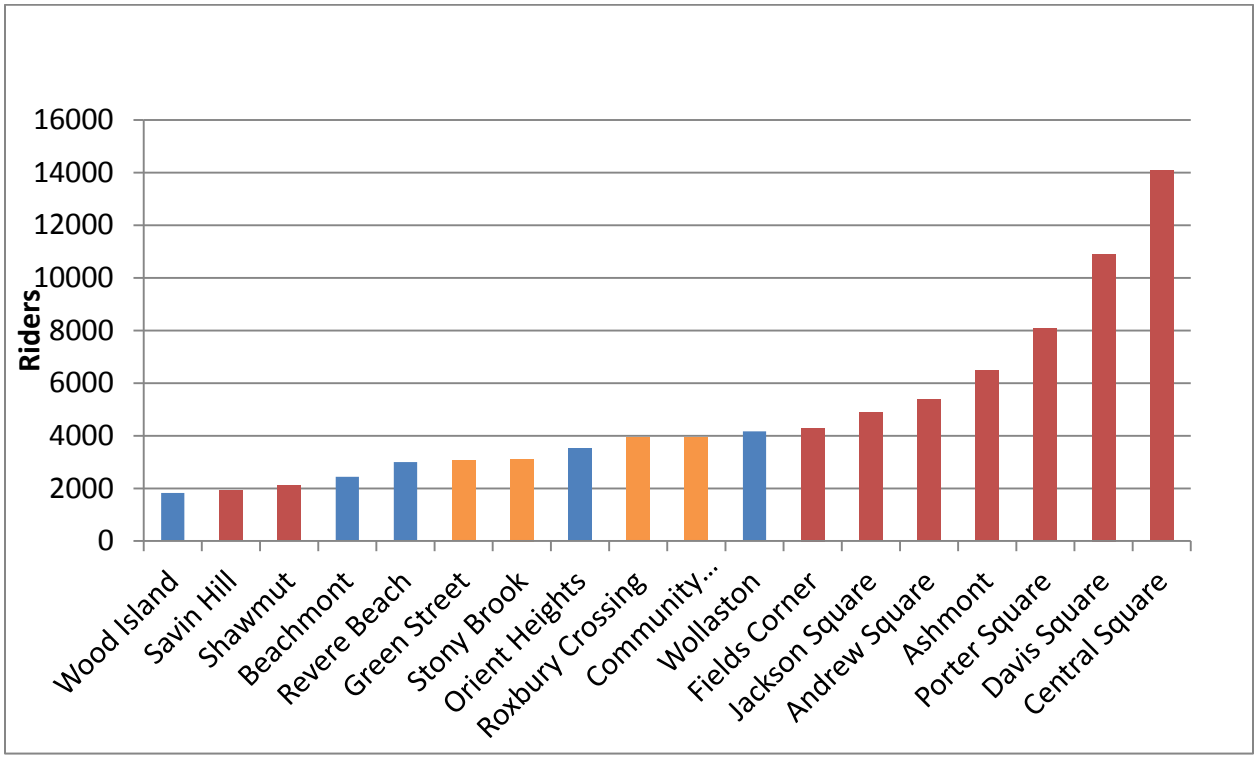
From this initial list of 18 stations, six were chosen from a variety of factors, including: ridership patterns (morning clump versus even spread), current land use zoning regulations (ranging from very dense residential to stations with a majority of zoning classified as open industrial or institutional and recreational), and InfoGroup and Foursquare data (which sites were most popular, outperformed or underperformed expectations based on population, etc.). The six stations ultimately selected were Central Square, Community College, Davis Square, Fields Corner, Orient Heights, and Revere Beach.

6 Case Study Stations	Subway Line
Central Square	Red
Community College	Orange
Davis	Red
Fields Corner	Red
Orient Heights	Blue
Revere Beach	Blue

Ridership

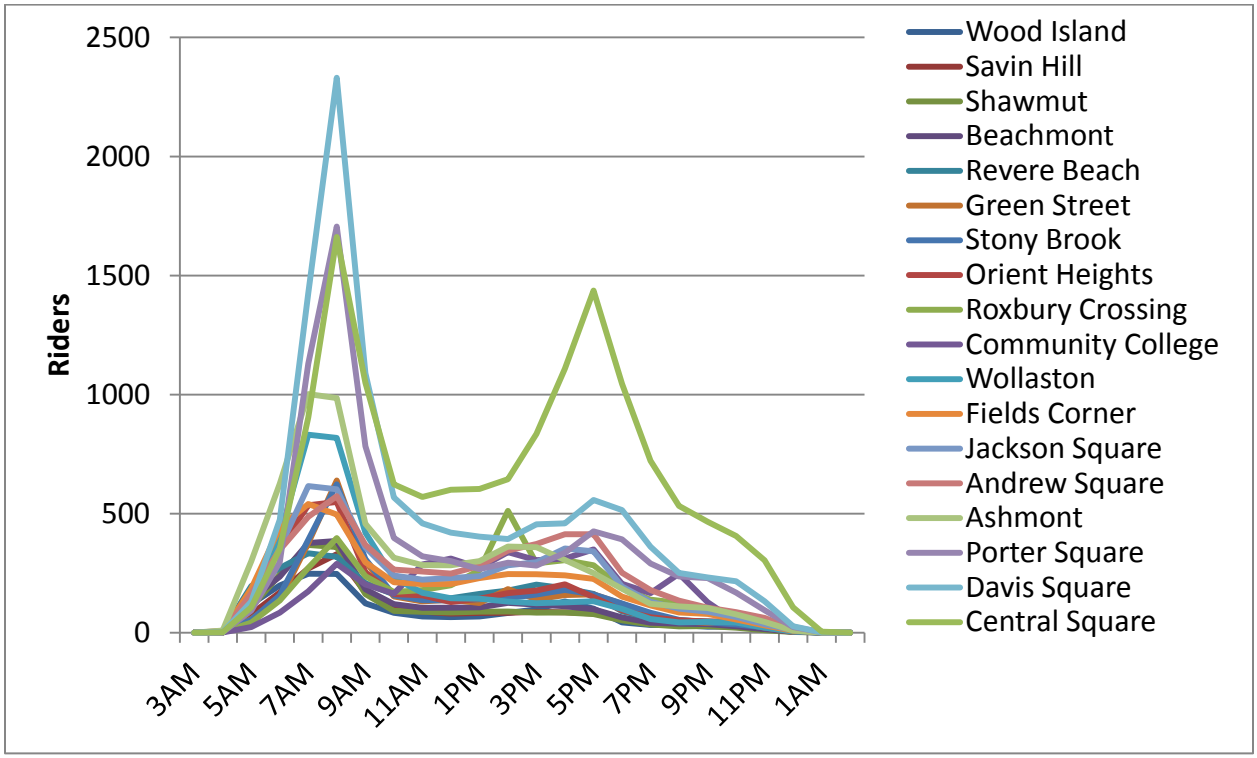
Data: MBTA Ridership Database, September 2009 to September 2010.

Weekday Average Ridership

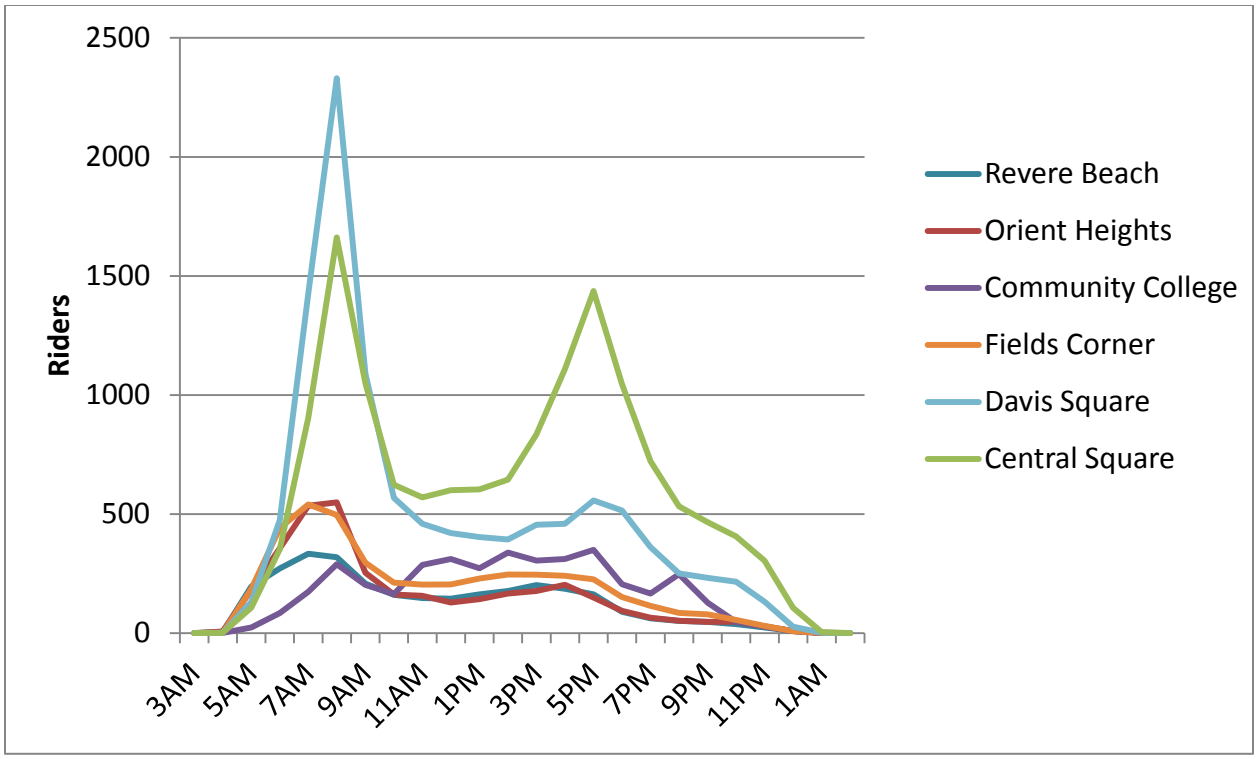


Analysis of weekday average ridership showed general trends among the subway lines. The red line overall had higher ridership than the orange and blue lines. Central Square, Davis Square, and Porter Square stood out as individual stations with notably higher ridership. From the six stations selected for case study analysis, Revere Beach has the lowest ridership with an average of 3,000 weekday riders and Central Square has the highest ridership with an average of 14,000 weekday riders.

Hourly Weekday Ridership

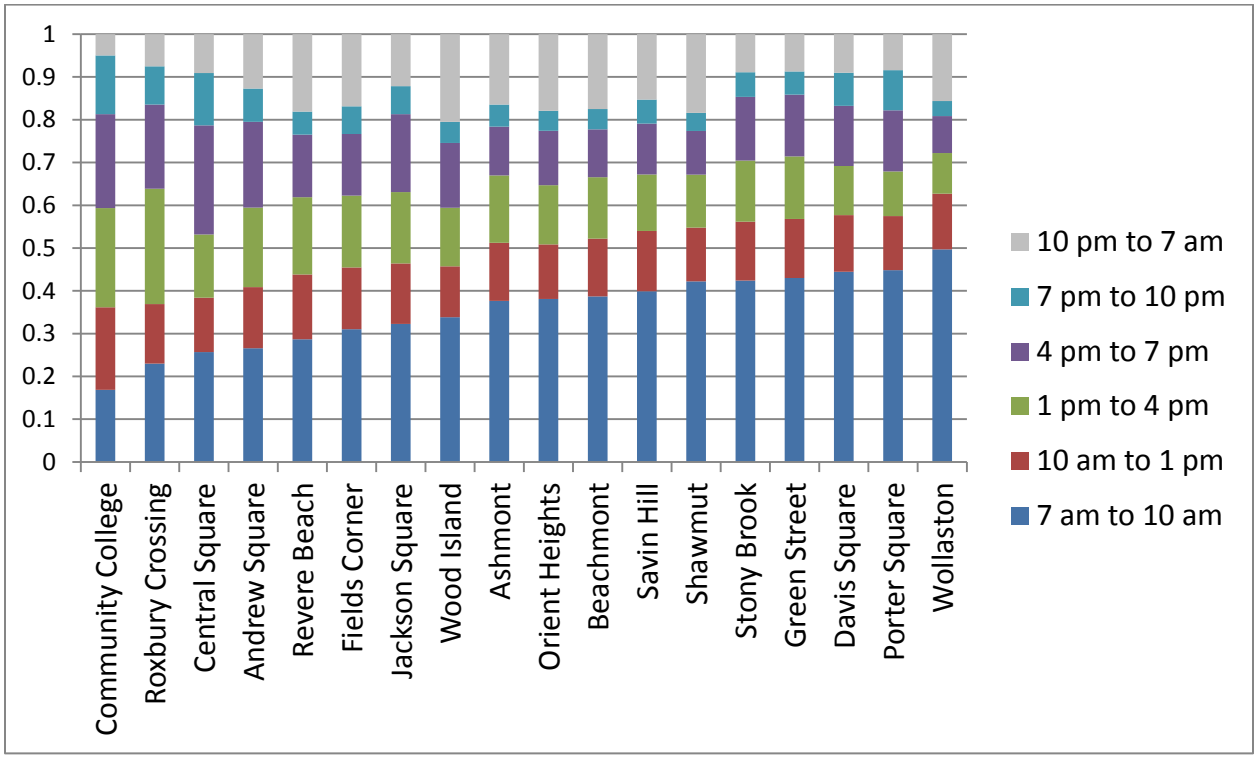


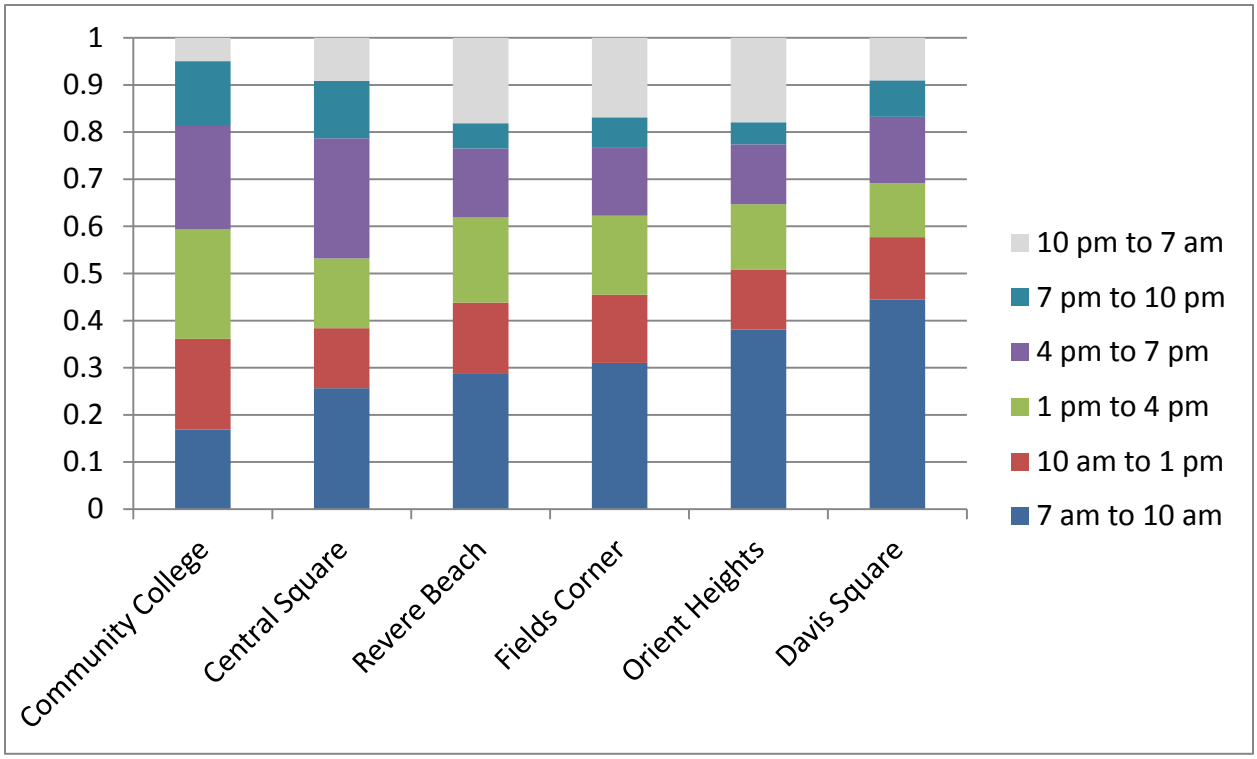
Analysis of hourly weekday ridership further illuminated differences between stations. Davis has the highest morning ridership, representative of its highly residential character. This analysis also gave the first indication that Central Square functions differently from the other eighteen stations. In the chart, Central has almost equal morning and evening peaks. While morning peaks imply that a nearby residential population is leaving for work, evening peaks imply that workers are commuting home after businesses close. Central Square supports residents and commuting workers to an extent that none of the other eighteen stations does.



Interesting characteristics are noted when looking at the six case study stations. Community College, while having lower gross ridership, has relatively even ridership throughout the day, possibly indicating the station has an attraction which draws riders even during off-peak hours. In this instance, the college located there likely attracts commuting students regularly over the course of the day.

Time of Day Ridership Proportions

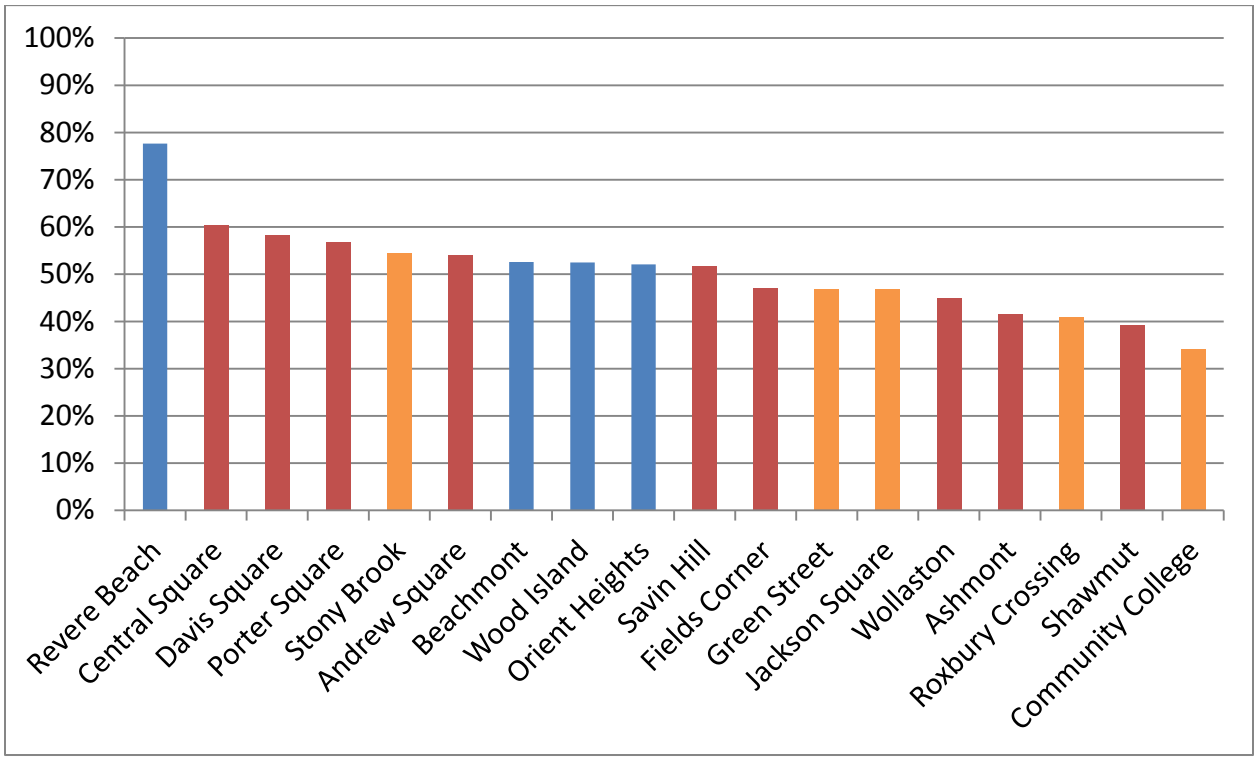


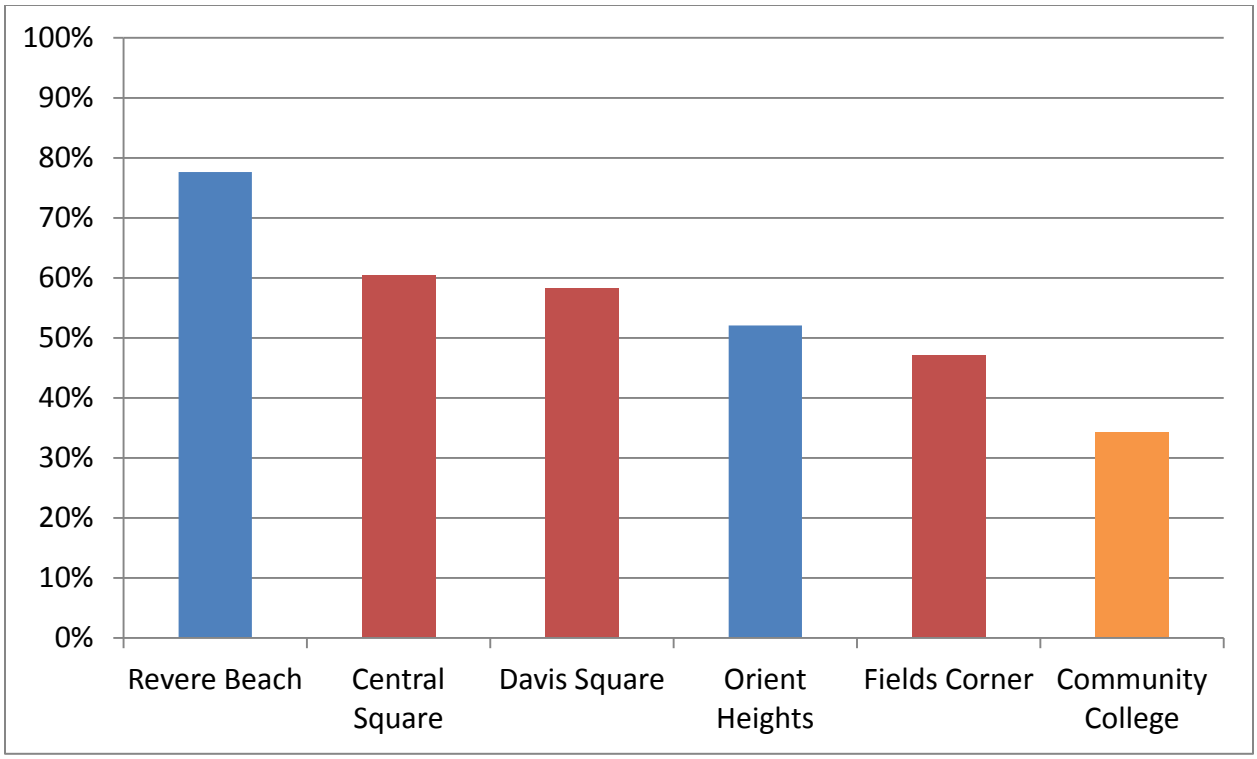


The time of day ridership proportion again shows Community College fairly even ridership load from 10 am to 7 pm. Davis Square has notably high morning ridership and Central clearly has even ridership from 7 am to 10 am and from 4 pm to 7pm.

Weekend Ridership as a Percentage of Weekday Ridership

Besides hourly comparisons, it was important to evaluate how stations performed on weekends versus weekdays. The following charts show average weekend daily ridership as a percentage of average weekday daily ridership for each of the 18 neighborhood subway stations.





Revere Beach stands out as having unusually high weekend ridership, with close to 80%. Community College has the lowest weekend ridership percentage. Both of these speak to the power of destination activities. Revere Beach has a beach which attracts weekend visitors. Community College on the other hand attracts students who do not go to the school on the weekends.

Urban Form

Data: MassGIS & U.S. Census 2010

Data & Methodology

To get a sense of the urban form surrounding a station we calculated indicators to measure availability of sidewalks, density of intersections, and street width, as well as comparing ridership to residential population and the distribution of land uses, all within an 800 meter radius of the station. For data we used Mass GIS roads and land use layers and aggregated Census 2010 residential data (put together on a 250 meter grid by last year's class). Street and sidewalk areas were calculated in GIS and then summed to get relative percentages (since all stations were being compared with the same surrounding area size). We present these relative percentages as well as a ratio of sidewalk to road area. Intersections of the roads layer were calculated using the Intersect Spatial Analysis tool and then subsequently the Spatial Dissolve tool to consolidate all intersections with the same X,Y coordinates. Mass GIS land use categories were aggregated to create seven simple categories – Open Natural, Open Industrial, Industrial, Residential, Commercial, Recreation/Institutional, and Water. Most of these categories are self-explanatory, although it should be noted that Open Industrial includes transitional, junk yards, transportation, and waste disposal, and Recreation/Institutional includes some open space land uses (like golf courses) as well as built land uses (schools, stadiums, churches, public buildings). Finally, in the maps provided of the area surrounding the station the roads are shown to emphasize whether they function as pathways or barriers. Highways and major roads (including Mass Ave which can be both path and barrier) are shown in orange and yellow respectively. Other roads are shown in white with diminishing thickness as they increase in class (ie. become more and more low-traffic and narrow).

Central

Compared to our other stations Central is really an outlier in terms of its incredibly and consistently high ridership and much larger (by at least 10,000 people) residential population. Although it is predominantly residential in terms of the surrounding land use area, it has a very high amount of commercial land as well. It also sets the high boundary for our urban form indicators (% of sidewalks, roads, and # of intersections).

Davis

Although Davis does not have as large of a population as Central it has the highest percentage of its population accounted for in morning weekday ridership, indicating the large number of morning commuters. It does not have as much commercial land use as Central but it still roughly equivalent (if not higher) urban form indicators.

Fields Corner

Fields Corner has the second largest residential population of our six exemplar stations, yet it has a much lower percentage of morning weekday ridership. It is interesting that although urban form indicators are still relatively high only 9% of land use is devoted to commercial uses whereas there is 14% for open industrial. Perhaps this shows some potential for commercial build out? We suspect that many of the open industrial land uses could be easily rezoned to encourage commercial development.

Orient Heights

Orient Heights is interesting because it has very distributed land use patterns. There is a lot of open industrial and institution/recreation land use, while there is very little commercial land. Also the urban form indicators are very low – only 90 intersections! – and looking at the map it is clear that some of the major roads here act as barriers to the station. Still, there is a high percentage of morning weekday ridership captured in the residential population.

Revere Beach

Revere Beach is another curious station because of the high amounts of open industrial land. Again, while it does not have the high urban form indicators or amounts of commercial land, the attraction of the beach and the consistent ridership seems to indicate potential for investment, especially if the open industrial land were to be developed for residential and commercial purposes. [Maybe too much on the recommendation side?]

Community College

In terms of its urban form Community College is a disastrous site. The I-93 highway is the epitome of a road acting as a barrier, completely transecting the area around the station so the southwestern half of the area is barely accessible. This disproportionate amount of infrastructure is shown in the majority of land use being categorized as open industrial (in this case all the highway, underpasses, and rail yard).

Destination Activities

Data: InfoGroup and Foursquare Databases

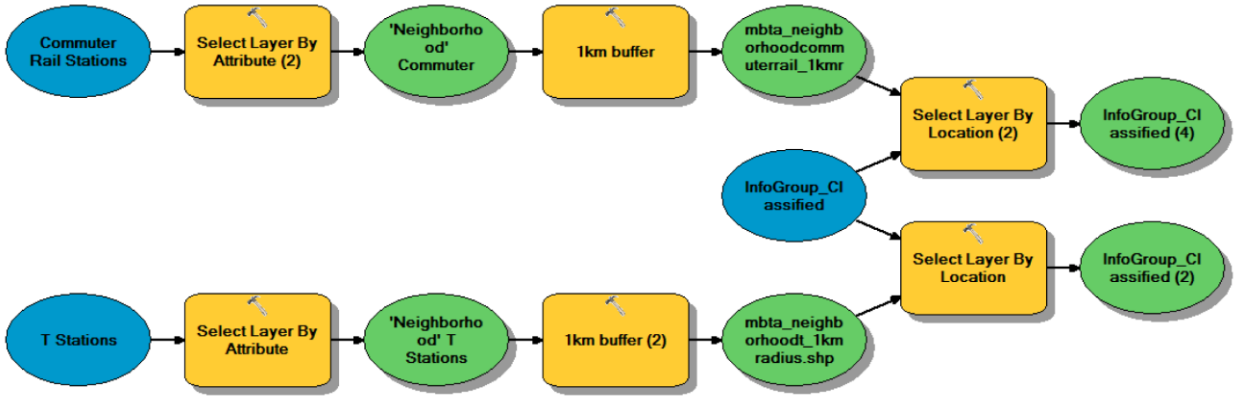
InfoGroup Data and SIC-Based Classification Scheme

The InfoGroup data is a commercial, for-profit database of local businesses and associated data collected through phone survey and updated annually. Through the Metropolitan Area Planning Council (MAPC), we had access to all businesses in the dataset within Massachusetts; these businesses constituted about 283,000 data points. Each business in the dataset has the following information:

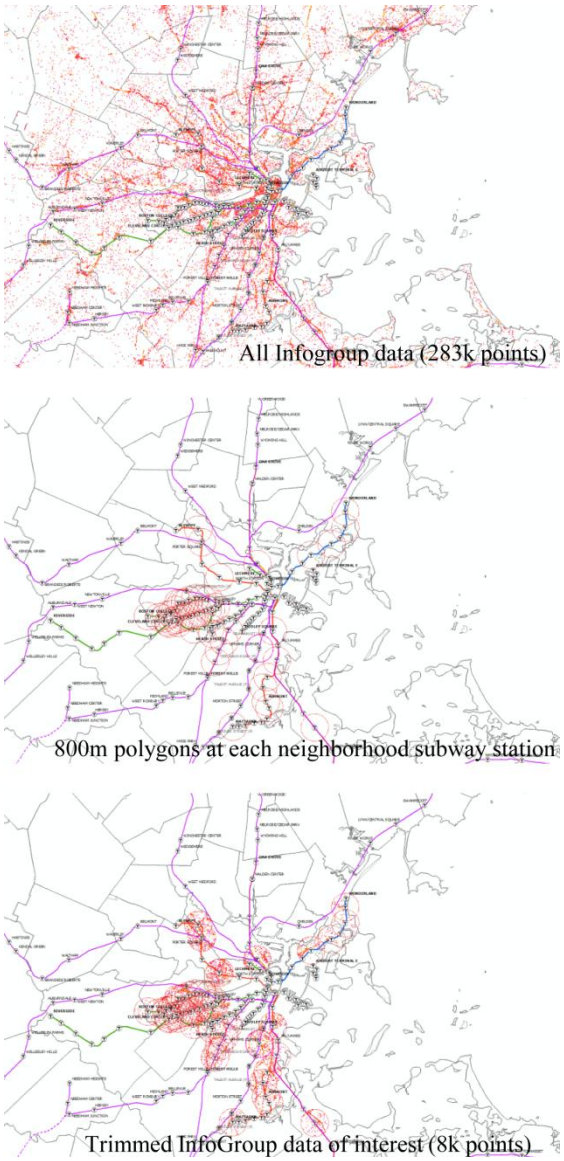
- spatial latitude and longitude coordinate data,
- business name,
- address,
- business type by SIC and NAICs codes,
- number of employees and
- sales volume

Among many other fields of data; however, in our analysis we found the above most useful. More information on the InfoGroup database is available in the white paper "Enhanced Business and Residential Data: The Importance of Coverage, Accuracy and Recency for GIS Data Sets" from InfoGroup (2010).

The InfoGroup data was loaded into ArcGIS and geolocated by its latitude and longitude coordinates. A model builder script was created to filter all businesses within 800m of our identified stations. 6,465 businesses were within this spatial query.



Model Builder Script to Filter Businesses Within A 'Crow-Flies' Distance of Stations



Process of Spatially Filtering InfoGroup Data

Using the Standard Industrial Classification (SIC) codes from the InfoGroup data, a classification system of the 1005 SIC codes was created of 34 types and subtypes, shown in the table below. Further each SIC code was given a destination flag indicating whether it is likely a business that would be visited by people.

Type	Sub-Type	Count	Count Destination	Percent Destinations
Agriculture and Farming		50	0	0%
Building Construction		26	0	0%
Communications		7	0	0%
Finance		34	9	26%
Forestry, Fishing and Hunting		8	0	0%
Insurance		9	0	0%

Manufacturing		459	0	0%
Mining		31	0	0%
Public Administration		28	0	0%
Real Estate		10	0	0%
Services		7	0	0%
Services	Art	2	2	100%
Services	Automotive	14	0	0%
Services	Business	32	0	0%
Services	Education	8	8	100%
Services	Engineering	13	0	0%
Services	Health	20	8	40%
Services	Organizations	7	0	0%
Services	Personal Use	15	12	80%
Services	Recreation	12	0	0%
Services	Repair	8	0	0%
Services	Social	5	1	20%
Services	Video	7	3	43%
Trade	Automotive	8	0	0%
Trade	Clothing	7	7	100%
Trade	Eating and Drinking	2	2	100%
Trade	Furnishings	9	9	100%
Trade	General	3	3	100%
Trade	Grocery	7	7	100%
Trade	Hardware	5	5	100%
Trade	Retail	23	23	100%
Trade	Wholesale	69	20	29%
Transportation		34	0	0%
Transportation	Local Transportation	12	0	0%
Utilities		14	0	0%

Foursquare

Foursquare is a location-based social networking site where users 'check-in' to destinations they are at with a cell phone application. Foursquare data publicly available through their API which we utilized includes:

- business name,
- primary category,
- check-in count,
- users count and
- tip (comment) count.

Using the Foursquare public API, InfoGroup business name and latitude and longitude data, businesses in the InfoGroup dataset were matched with the above Foursquare data. (The Foursquare API provides a search method, `match`, designed to match against existing destinations based on a fairly precise location and a fuzzy name string. This is what we used.) Initially 2,554 Foursquare matches were made

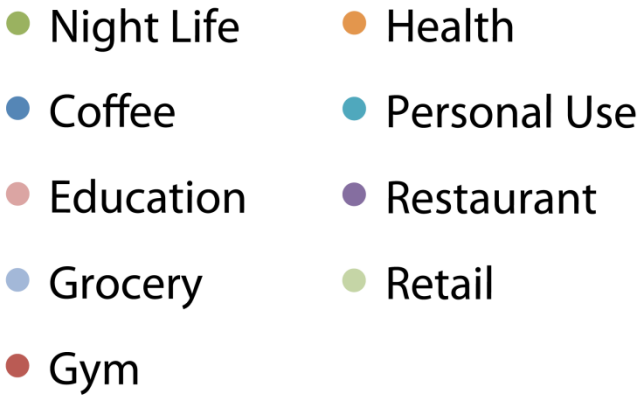
to the 6,465 InfoGroup businesses (39.5%), but upon further investigation for quality and removal of duplicates it was found that only 2,242 (34.7%) matches were of good quality.

As Foursquare users find venues they check-in to desirable enough to share with their friends, we believe that the matched data constitutes a good representation of 'destinations.' The excluded InfoGroup businesses which were not matched by Foursquare venues are likely primarily places where people might work but are not destinations.

The Foursquare matching API may return multiple potential matches for a given business name, location pair. 364 (14.2%) of the matched InfoGroup business entries matched multiple Foursquare venues. These predominantly corresponded to multiple destinations with similar names, near or in the same physical location, e.g. the Museum of Science, and the Museum of Science Café. In the InfoGroup database, there would commonly be only one business entry for such a location, while Foursquare's venue database provides many fine-grained entries created by users, down to the level of individual classrooms in a college, for instance. These multiple matches had to be manually filtered down to the most meaningful single match using a filtering table and an additional view layer in our destination database. This avoids over- counting a single business, but it also risks under-counting meaningful venues in Foursquare. A more ideal scheme might coalesce multiple Foursquare venues corresponding to the same InfoGroup business into a single aggregate entry, but this would require substantially more manual effort and discretion.

Finally, Foursquare venues matched multiple times by different InfoGroup businesses (292 in all) had to be filtered. Most of these were corrected during the manual matching of multiply matched InfoGroup businesses, but some remained after. For these, we focused on removing redundant matches for those that had high check-in counts and would therefore skew our data significantly.

Our Classification Scheme

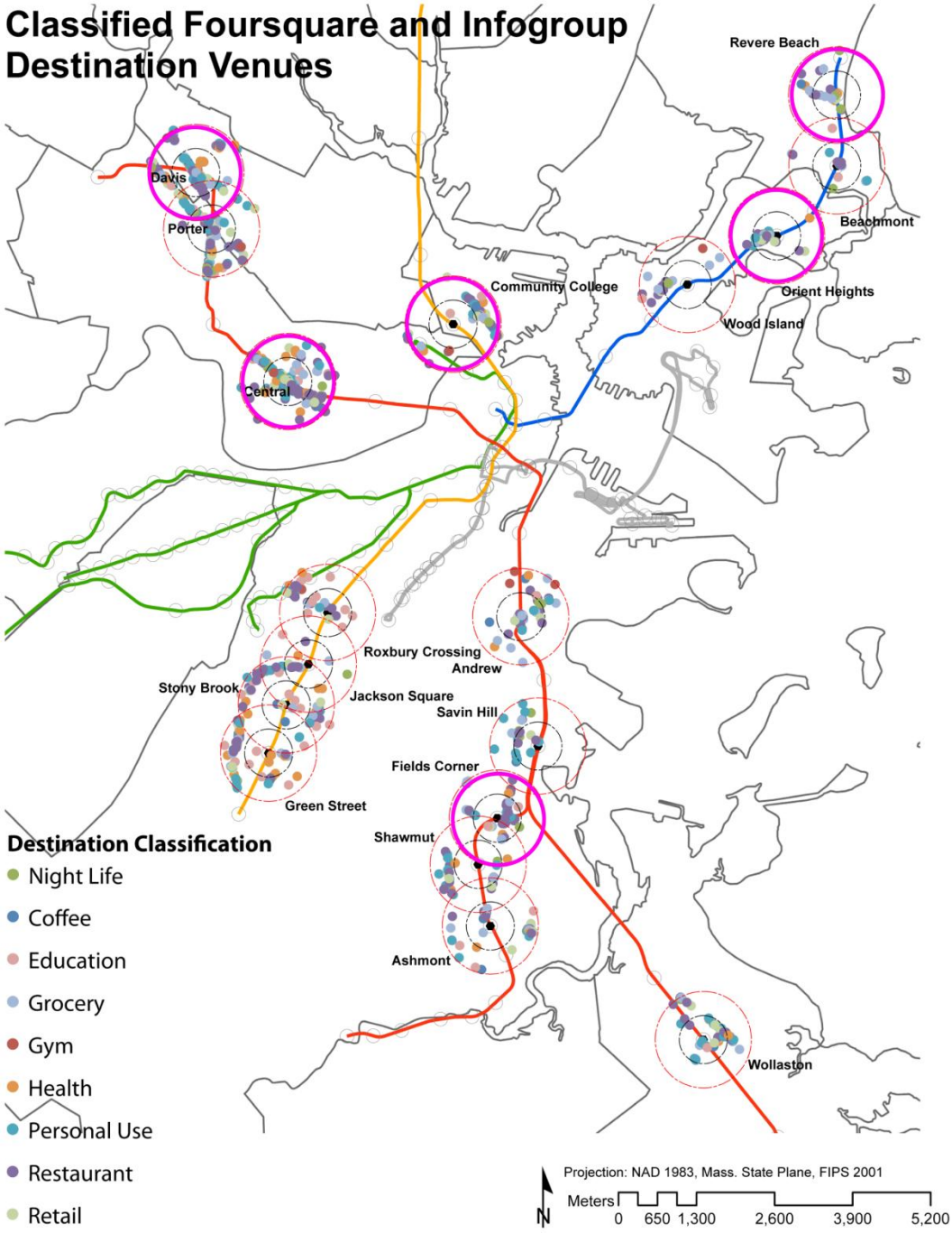


Our Nine Category Custom Classification Scheme

The matched Foursquare and InfoGroup venues were categorized based on observed trends in the data through running queries with 'group by' statements by Foursquare primary category (jrk.mapc_foursquare_venues, primary_category) and custom types and subtypes based on the SIC

code (jrk.mapc_infogroup_venues, type & subtype). The categorizations and their definitions are explained below,

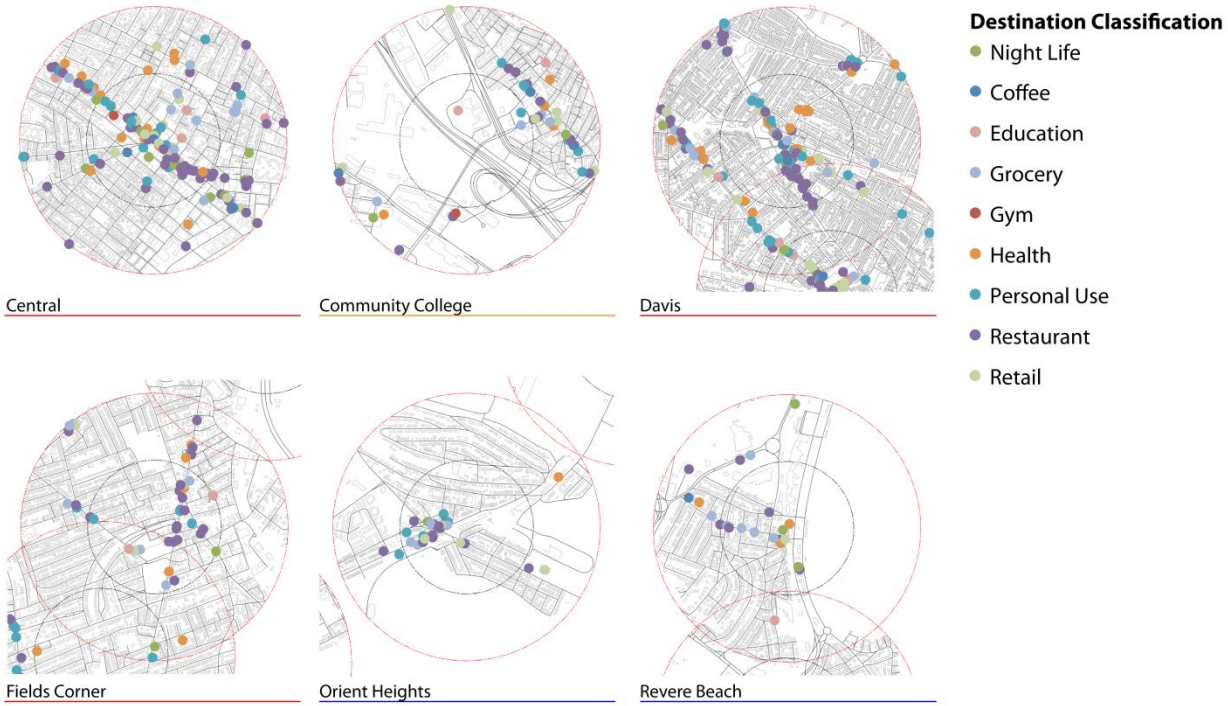
- Night Life
  primary_category: cocktail, bar, sports bar, pub, lounge or night club
- Coffee
  primary_category: coffee shop
- Education
  subtype: education
- Grocery
  subtype: grocery
  NOT primary_category: coffee shop
- Gym
  primary_category: gym
- Health
  subtype: health (doctor's offices, hospitals)
- Personal Use
  subtype: personal use (laundromats, barber shops, dry cleaners, beauty shops)
- Restaurant
  subtype: eating & drinking
  NOT 'Coffee' or 'Night Life'
- Retail
  subtype: retail (anything not wholesale, consumer sales)



Resulting Destination Information

Using the above methods, detailed maps of our primary six stations were created.

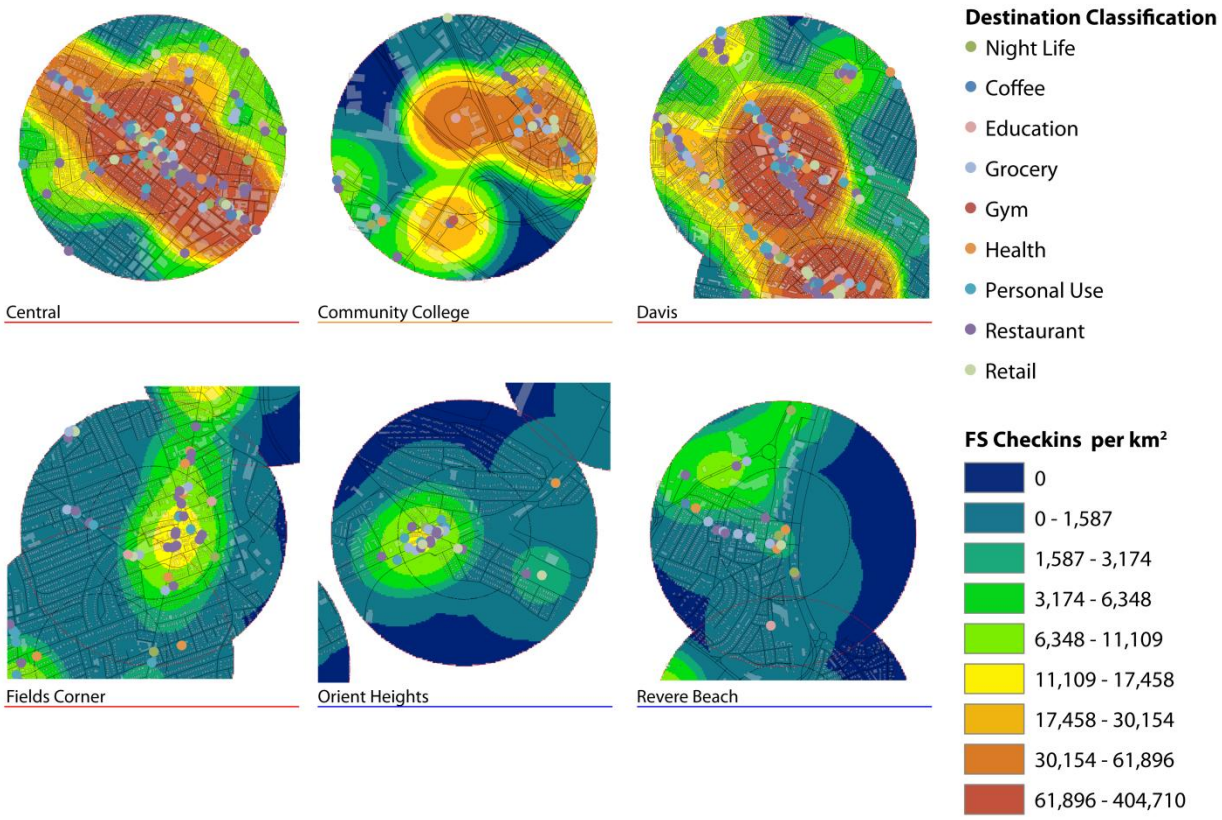
The figure 'Classified Destination Maps' below shows the classified destinations plotted on top of 800m radii surrounding each station. The smaller radius in the center of each map is 400m, for the purposes of scale.



Classified Destination Maps (InfoGroup & Foursquare Matched)

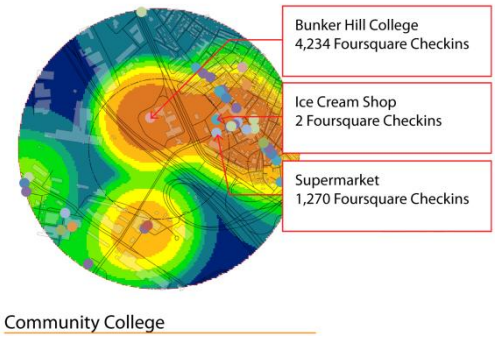
From these maps it can be observed that in general destinations tend to be clustered along streets. Central and Davis, the two most dense in terms of numbers of destinations, also have these commercial streets running directly through the station center. Fields Corner, Orient Heights and Revere Beach, the next tier in terms of density of destinations, also see an increased clustering near stations. Fields Corner seems to have two commercial streets intersecting near the station. Orient Heights and Revere Beach have clusters of destinations near the stations themselves. Community College is the clear outlier here; it has only three destinations within the 400m radius of the station. The nearest observable commercial street is in the 400-800m distance area to the North of the station. It is also worth noting that the destinations South of Community College are rendered inaccessible by highway 93. One other interesting observation is that in the area around Central, several destinations seem to be diffusely located in the neighborhood and not associated with any specific commercial street.

The next step is to look at popularity of these destinations measured by the density of Foursquare check-ins, shown below using a 400m linear falloff from each destination location.



Foursquare Check-In Density

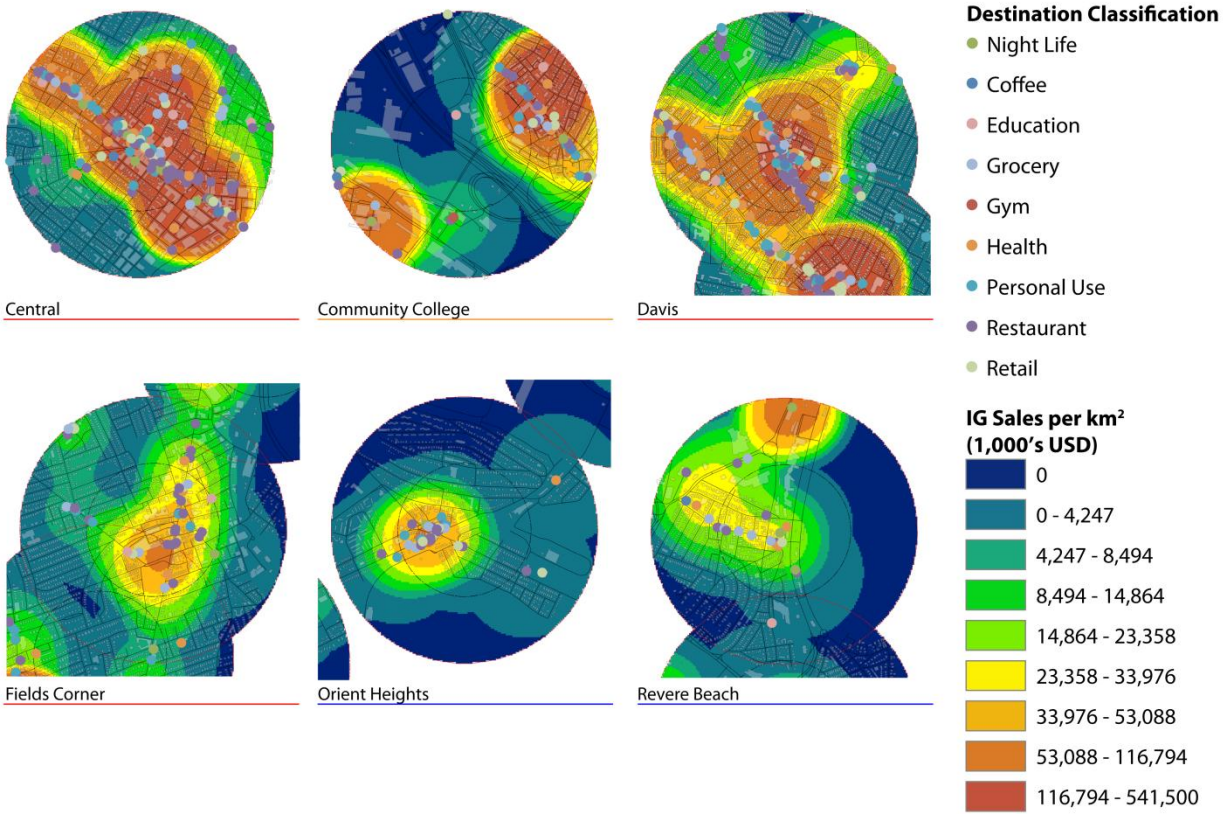
The check-in density maps again show the relative dominance of Central and Davis in popularity with Foursquare users, as one might expect from ridership trends and the number of destinations at those statiosn. Fields Corner and Orient Heights do moderately well near their stations; however, Revere Beach does worse than expected. As Revere Beach and Orient Heights both have beaches near their stations, it is easy to wonder why this is. Also, Revere Beach's most popular locations are to the north, along a very high traffic road. This speaks to the potential that car traffic at Revere Beach dominates public transportation currently.



Example of Destination Information, Community College

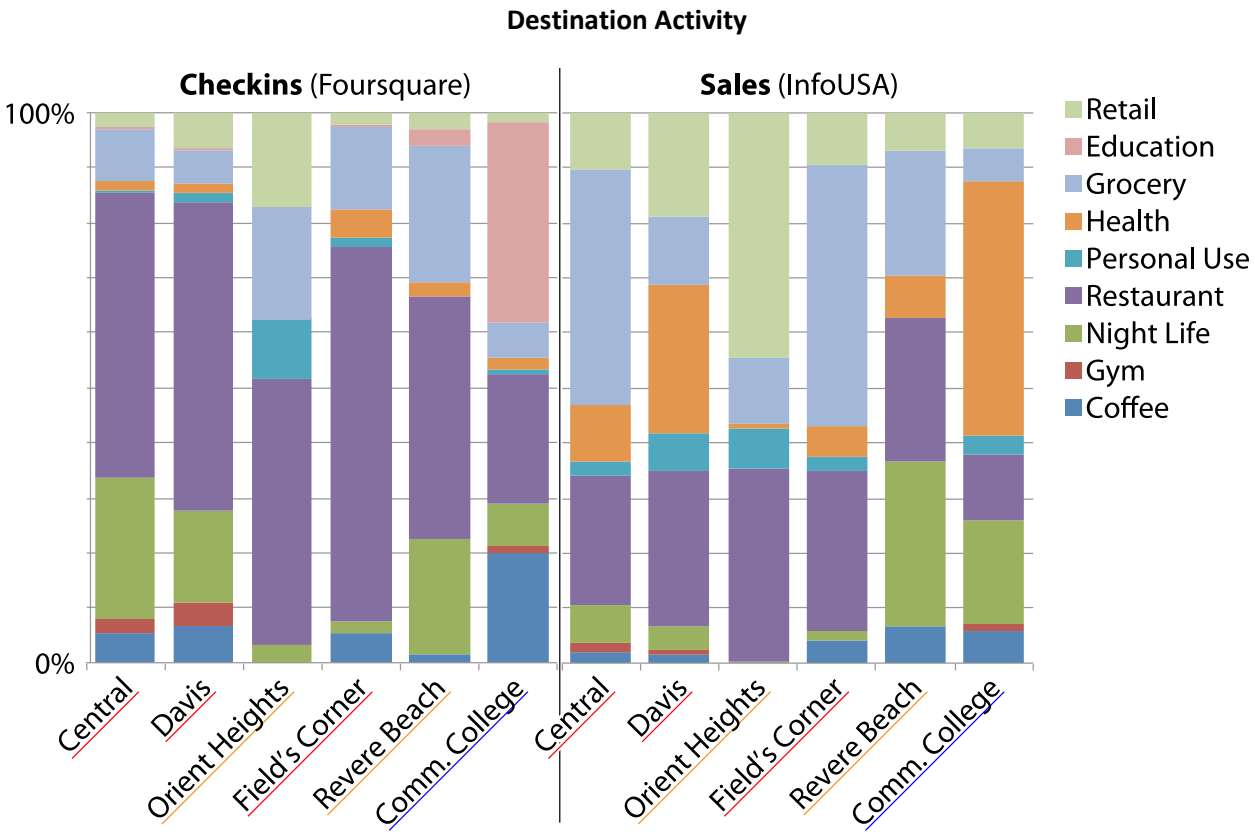
Finally, we notice that Community College has a hot spot of check-ins around the station, although before we noted that there were only three venues within 400m of it. This can be explained by two of the venues having relatively high check-in counts: the college itself and a supermarket.

The matched Foursquare and Infogroup destination data may also be looked at in terms of commercial popularity by sales in dollars. This data is necessarily imperfect, as we would like to plot it by number of sales or normalize it by the average value of a sale in each destination; however, this was not possible with our imperfect knowledge.



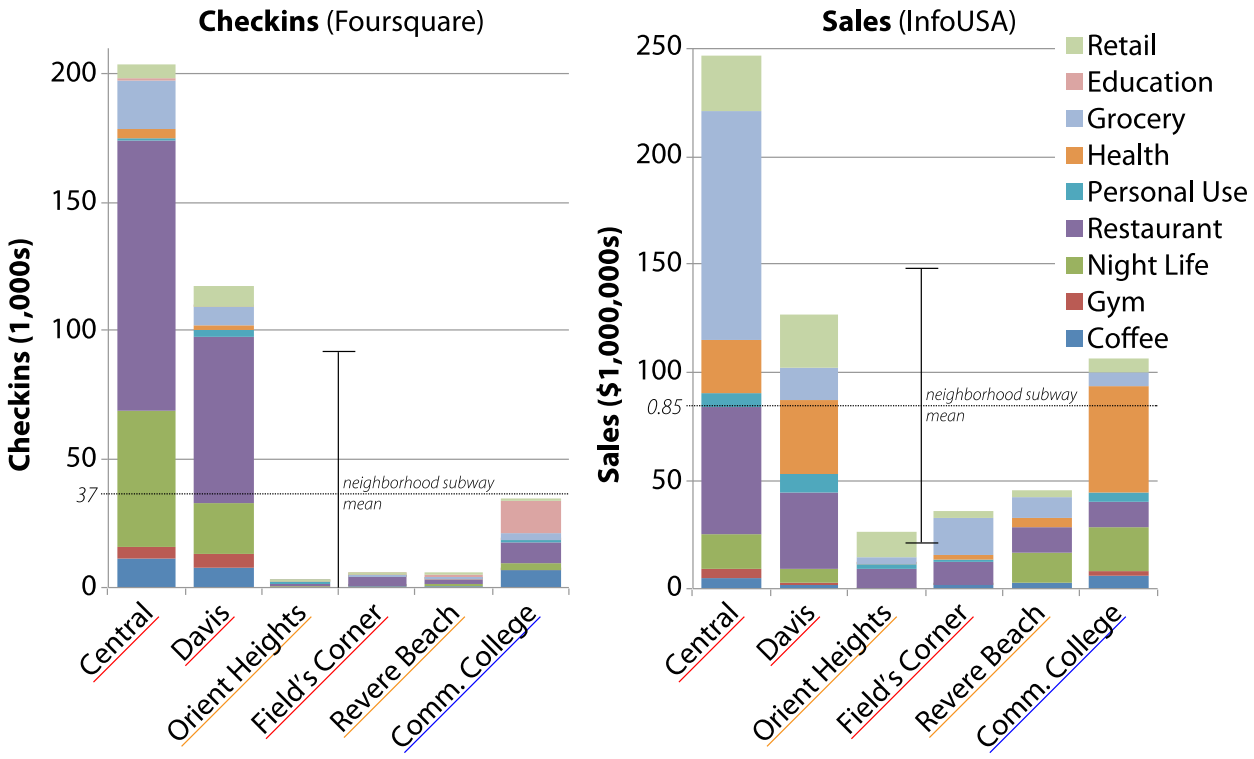
Infogroup Sales Density

Observing sales density, again it is seen that Central and Davis are dominant compared to other stations. Fields Corner, Orient Heights, and Revere Beach all do somewhat better in terms of sales comparatively to their check-ins performance. This seems to suggest that there is a demographic not being represented in the Foursquare data or that these businesses are being used more by locals who would not check-in on Foursquare. This latter explanation suggests that there might also be potential to increase the 'destination-ness' of these stations. Finally, Community College has a complete reversal of relative density between check-ins and sales. This makes sense as a college is likely not to have any saleable goods, but it is popular with the student demographic.



Proportion of destination checkins and sales by category.

Here we have broken down the relative fraction of checkin and sales activity by destination category. Checkins are heavily weighted towards restaurants and night life. Education is highly represented at Community College, where all such checkins are directly at the Bunker Hill College venue. Sales, meanwhile, are dominated by grocery, general retail, restaurants, and, in some cases, health.



Absolute checkins and business sales by category. Mean and error bars mark the mean +/- 1 standard deviation for total destination checkins and sales across the entire cohort of 18 *neighborhood subway* stations studied.

Looking at absolute checkins and sales, there is a dramatic difference in scale among the stations. Checkins are heavily biased towards Central, Davis, and Community College. This suggests a bias you might expect in a smartphone-based social network, towards the young and/or affluent, and towards more urban settings.

Sales are more in line with ridership activity at the stations. The enormously high health sales at Community College correspond to the MGH Charlestown hospital, which is captured in our 800m window. At Central and Davis, meanwhile, high health sales correspond primarily to dentists—more retail storefront-type destinations. The largest volume of grocery sales is explained by just a handful of stores. At Central, 71% is one Star Market and one (small) Whole Foods. The restaurant, retail, and night life categories, meanwhile, are smoothly distributed across many venues at each of the stations.

Looking at individual stations, it is interesting to compare Central to Davis. They have similarly high sales and checkins, roughly proportional to their relative population and aggregate ridership. But recalling their ridership profiles, it is clear that they reach this success in very different ways. Central could nearly be categorized as an urban core station. It has high ridership at all hours, and on weekends, indicating a rich mix of residents, workers, and visitors to its businesses. Davis, meanwhile, is dominated by weekday morning tag-ons and a heavily residential zoning mix, indicating that its activity corresponds primarily to residents consuming its rich array of local businesses. Other stations fall far below in activity, roughly in order of ridership. Even the number of destinations falls far below Central and Davis, suggesting that these metrics are reasonably indicative of station quality.

Comparing Davis to Fields Corner—a superficially more similar station—both have similar populations (16.5k in Fields Corner, 15.2k in Davis), but Fields Corner has far less activity. This seems to match the zoning disparity between the two: while Davis is rich with businesses and commercial space, Fields Corner simply lacks business destinations—it is nearly all residential space.

Finally, Community College is an extreme outlier. It has high apparent activity, appearing third behind Central and Davis on both sales and checkins, with half the population of Davis. But digging deeper into the data paints a different picture. This measured activity is dominated by individual destinations in unusual categories—the college, the Museum of Science, MGH Charlestown, and the Royal Sonesta Hotel—most of which are near the periphery of our 800m radius of analysis, and several of which are separated from the station by the massive I-93 interchange which stops nearly all pedestrian flow through the area.

Conclusions

After looking at these three different kinds of data it is clear that understanding transit oriented development is not a straightforward task. Each kind of data on its own seems to indicate different stations to be winners and losers in terms of their TOD success. Take the example of Community College. In terms of ridership that station stood out for its consistently high numbers throughout the weekday. However urban form indicators speak to an insurmountable challenge in the form of the highway cutting through the site and cutting off the station to half of its nearby neighborhood. These stories seem to be at odds with each other until we account for the business and check-ins data. Then we begin to see a nuanced view of both the strengths and weaknesses of the station – the college functions like a strong destination, but the surrounding retail shows some room for improvement. Thus, our ultimate conclusion is the need to triangulate the success of a TOD site with a variety of data.