BUDT 730

Data, Models and Decisions

Lecture 01

Course Information & Introduction to DMD

Prof. Sujin Kim

Lecture 1

- The Course Overview & Logistics
- Introduction to Data Models and Decisions (CH1)

UMD's Early Contact Identification (ECI) in Classrooms

We must remain vigilant and committed to supporting each other, and do whatever we can to prevent the spread and further waves of COVID-19. To protect yourself and others, the Health Center is asking ALL students to adopt TWO critical and very simple practices:

1. At the beginning of each class session, please, **use your mobile device to "check-in" and scan the QR codes nearest your seating locati**on.
2. Do your best to remember and sit on the same seat or location (table or zone) while attending each session in the classroom.

More detailed information about ECI is available at the Division of IT website.

Instructor

Prof. Sujin Kim

- **E-mail:** kimsj22@umd.edu
- **In-Person Office Hours: Starting from Sep 8**

  Wednesdays 10:00-10:45 am, VMH 4335

  Thursdays 2:00-3:00 pm, VMH 3332

In-person office hours will be shared with all students who take BUDT 730, but the space is very limited (10). Therefore, the office hours will be restricted to those students who need direct help from Prof. Kim: some technical issues with software, some questions regarding lectures or any other matters relevant to this course. If you want to discuss some personal matters, please use these office hours.

Teaching Assistants

Jiangkun Xiong: Section 0501

- Email: jxiong02@umd.edu
- Zoom link: https://umd.zoom.us/j/5545909045

Qiuping Lin: Section 0502

- Email: qlin1@umd.edu
- Zoom link: https://umd.zoom.us/j/3803862662?pwd=VXJ3NTdhajhQMFN3bHIzdDY5NC9Edz09

Zhuxuan Xu: Section 0503

- Email: zhuxuan@umd.edu
- Zoom link: https://umd.zoom.us/j/9863890975

Jiaqi Huang: Section 0506

- Email: jhuang14@umd.edu
- Zoom link: https://umd.zoom.us/j/6256331676

Yizhe Chen: Section 0507

- Email: ychen345@umd.edu
- Zoom link: https://umd.zoom.us/j/8973770361

Schedule of Office Hours

	Wednesday	Thursday	Friday
9:00-10:00am			Jiangkun
10:00-11:00am	Prof. Kim (10-10:45 am, VMH 4335)		Jiangkun & Qiuping
11:00am-12:00pm			Jiaqi
1:00-2:00pm			
2:00-3:00pm		Prof. Kim (VMH 3332)	Yizhe
3:00-4:00pm	Jiaqi		Yizhe
4:00-5:00pm		Zhuxuan	
5:00-6:00pm		Qiuping	

All TAs' office hours will be held online.

If you have questions?

- If you have any questions about assignments, please attend any TA's office hour or send an email to your section TA. Try to send your questions by 6 pm on the due date so that your TA can have a time to answer them.
- If you have any question regarding grading (assignments and quizzes), please contact your section TA.
- If you want to discuss any other matters relevant to this course, please come to Prof. Kim's in-person office hours or email her.

Assignments

- **Assignments:**
  - There will be 8 assignments in total (6 individual and 2 team assignments).
  - You can find the schedule on Canvas: go to syllabus.
  - The course website must be used to submit individual assignments. Any assignments submitted via email will not be accepted.
  - All assignments will be due on Fridays at 11:59 pm.
  - A half-day grace period will be given to you (12 hours passed the deadline). You can submit your assignment during the grace period but with "late submission" mark. You should email your TA about the late submission with a clear explanation for submitting the assignment late. Otherwise, your assignment will be partially counted.
  - The submission will be closed after the grace period and any late homework after that time will not be accepted.

- **Quizzes**

  A number of quizzes will be given throughout the semester; online quizzes via Canvas. No make-up quiz will be given.

Course Grade

- Homework assignments: 30%
- Quizzes and in-class assignments: 10%
- **Midterm exam: 25% (Monday, Oct 18)**
- Final exam: 35% (TBA)

Please note that **no extra credit** will be given at any point in time throughout the semester.

Software

- This course focuses on the practical implementation of analytical techniques using computing technology that is widely used in the business world.
- We will use **Excel, Tableau**, **R, RStudio, and Palisade Decision Tools Suite.**
- All of this software will also be available on vSmith for free use.
- If you have any issues with software, please contact your section TA.

- We will learn R in our course (probably from Week 4 or 5); however, the following LinkedIn Learning course can give you a head start: Learning-R

Announcements: Quiz 1, Quiz 2 & …

- Quiz 1 has been posted on Canvas. It is due on Saturday, Sep 4.
  - Study the course syllabus
  - Install software packages or VMware to your PC/Mac
- Quiz 2 has been posted on Canvas. It is due on Saturday, Sep 4.
- IA1 is due on Friday, Sep 10, 11:59 pm
  - Take pride in your submissions!
    - Answer each question clearly and directly
    - Follow guidelines
  - Remember Academic Integrity!!
    - Consult with peers on general concepts and software, not on detailed process of completing the assignment
    - You must produce and submit your own original work!!
- Teams for TA1 (Due on Sep 17):
  - Go to Canvas and find your team for TA1 and contact your team members.
- No class on Monday & Tuesday next week. Please, study Tableau tutorials before you come to class on Wednesday.

Introduction to Business Analytics (Ch1)

What is DMD?

- Technically, this course should be called **Data, Models, and Decisions** (i.e., with commas)
- These components are distinct from each other -- you can have one without the other
- These are tools used in business analytics.

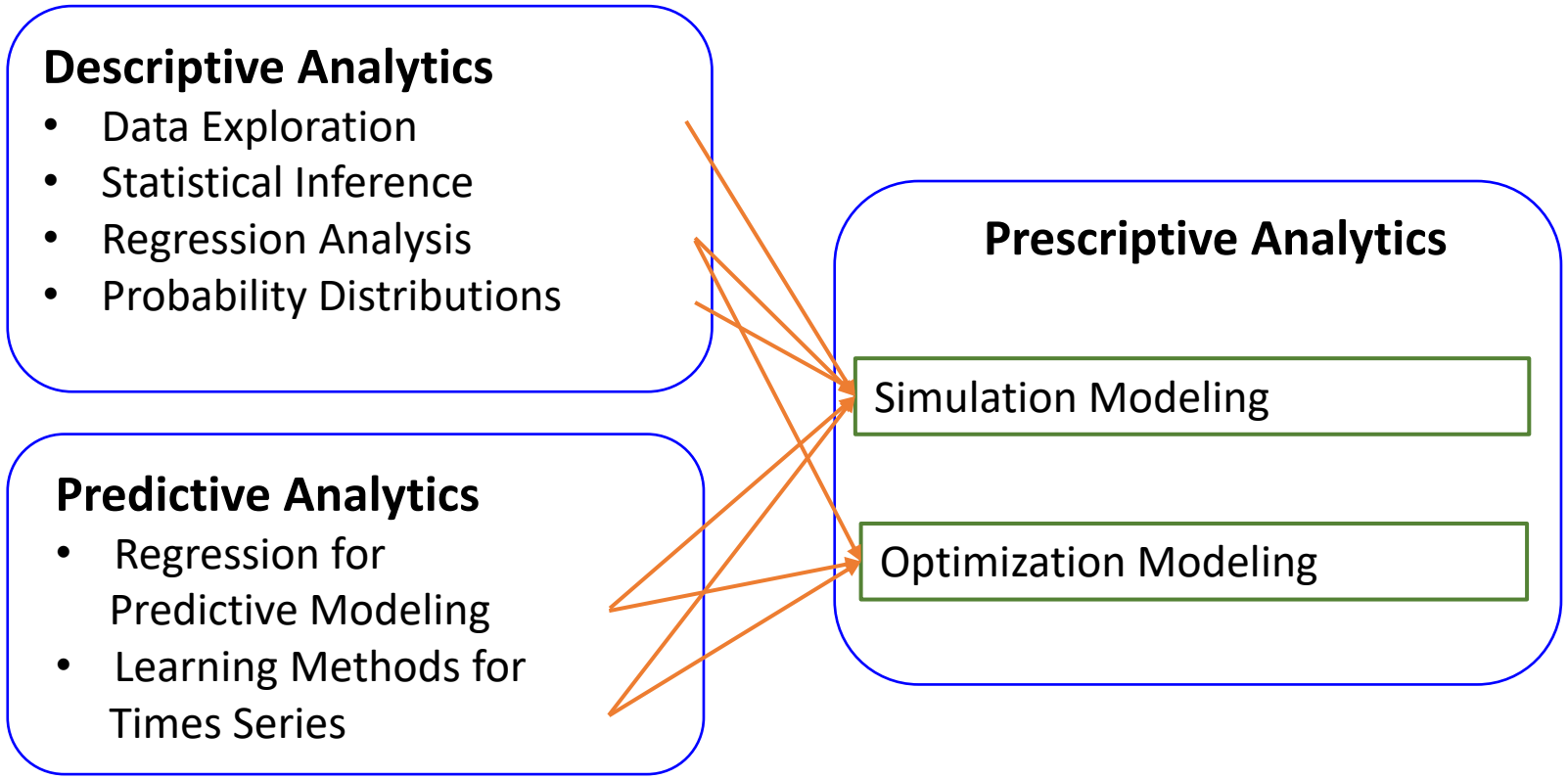
Business Analytics

- We live in the technology age
  - Possible to collect Big Data
    - A lot now, even more in the future
  - Provides the opportunity to analyze data and make decisions on the basis of quantitative analysis
- Business Analytics is the use of:
  - *Data,*
  - Statistical Analytics and
  - Quantitative methods
- To help managers and other stake holders gain insights and make better *decisions*.

Business Analytics

- Business Analytics may be divided into three areas:
  - ***Descriptive analytics*** focuses on gaining insights using historical data
  - ***Predictive analytics*** focuses on predicting unknown/missing/future data using statistical methods
  - ***Prescriptive analytics*** focus on decision recommendations using quantitative methods

Course Roadmap



Applications of Business Analytics

The application areas endless, examples include*:

Air traffic control, banking, canal operations, communications (broadband, broadcasting), consumer products, crowd control, express delivery, defense, education, financial (pension, investment, credit card,), fire protection, forestry, healthcare (treatment, operations, diagnosis, disease control), hotel management, energy production and distribution (coal, gas, electric, oil, nuclear), land use, manufacturing (electronics, food, paper, seeds…), marketing, printing, sanitation, security (airport, police), sports, social networks, tax collection, transportation (airline, highway, space,…), treasure hunting, waste management, water resources, quality, flow, flood), and weapons dismantlement

* Adapted from: https://www.informs.org/content/download/319043/3022847/file/2016_edelman_book.pdf

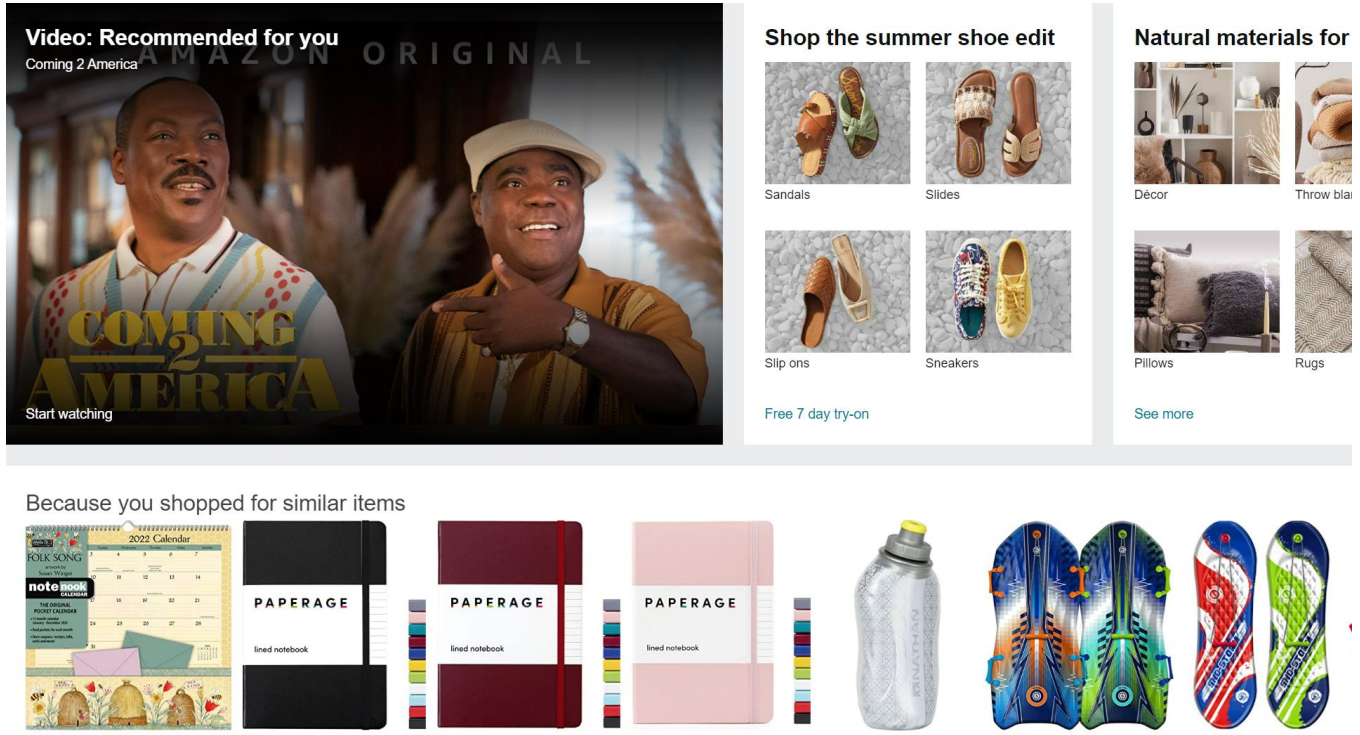
Business Analytics in the Real World

- Enlightened companies use business analytics to uncover the insights from the data AND acting on these insights
  - Amazon, Google, Netflix, UPS, Disney
- Others are creating business from their analytics talent base
  - IBM, Accenture, GE, many smaller firms
- Videos on Business Analytics
  - McKinsey: Making Data Analytics Work
  - Tom Davenport: Business Analytics Defined

amazon

- Which fulfillment center should be used for a particular order?
- How to dispatch delivery trucks?
- How to allocate agents to call centers?

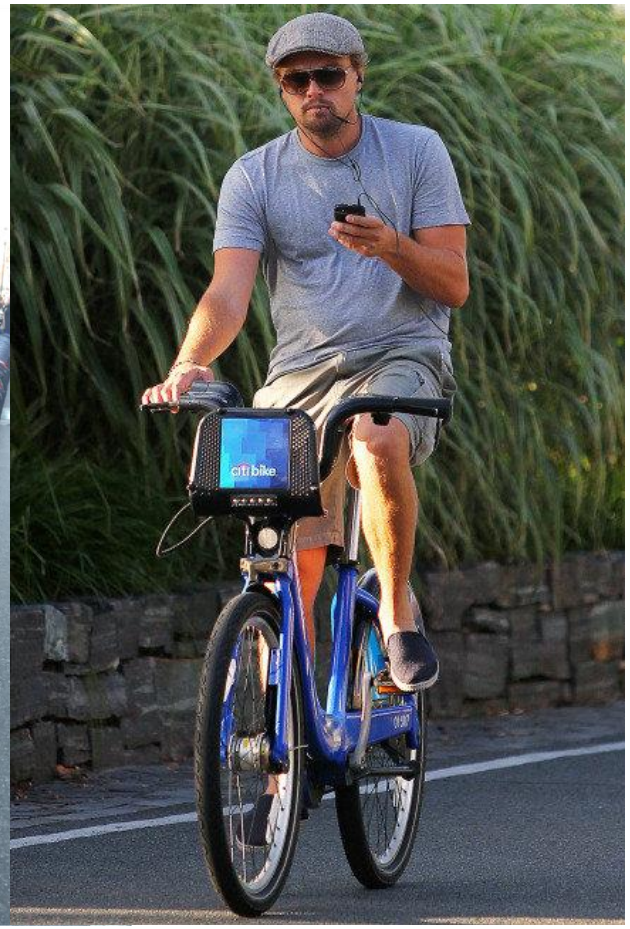
- How to dynamically change the price of each product?
- How to personalize product recommendations?
- Many more …



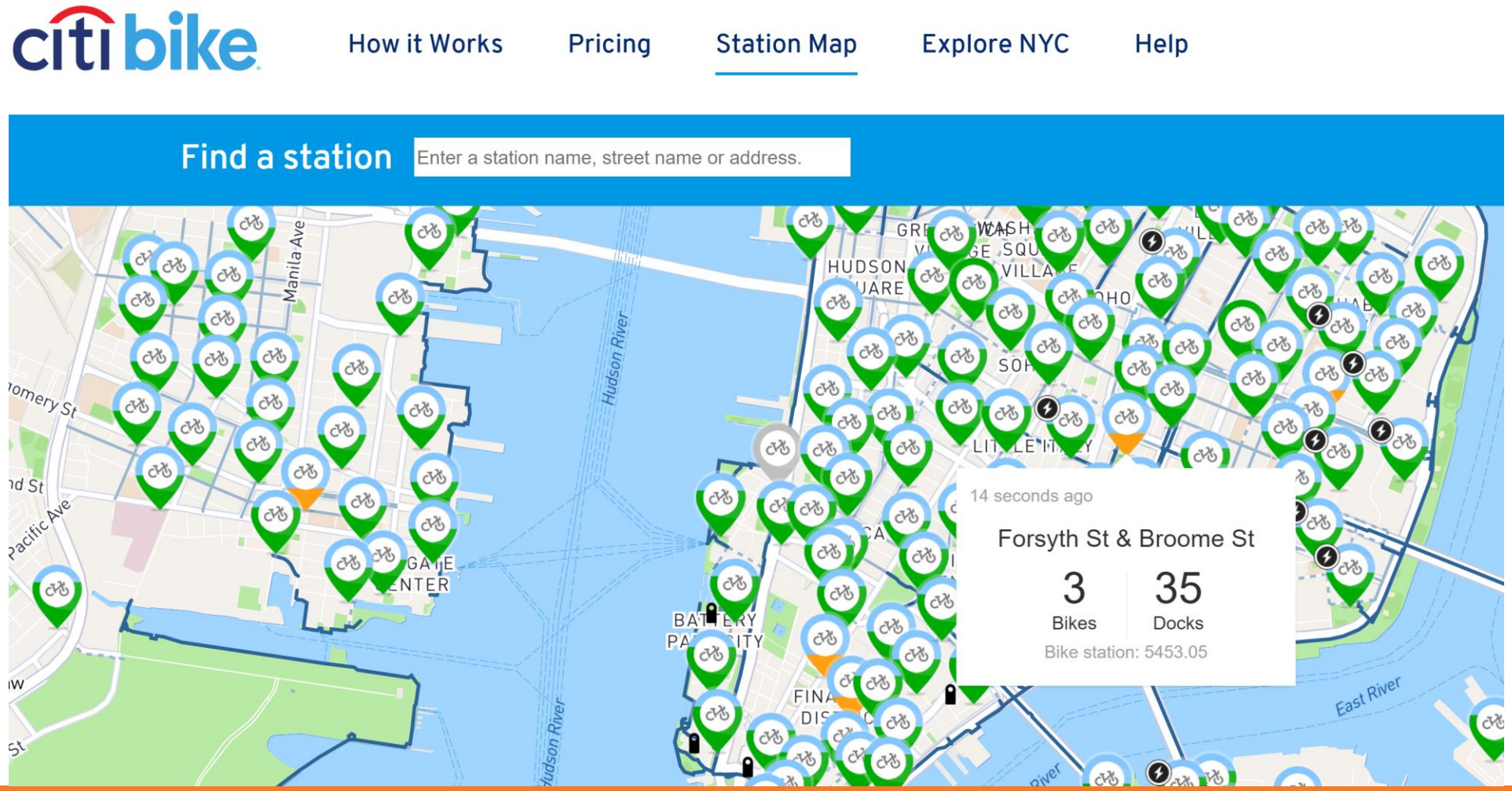
Example: Bike Sharing System

- Goal: To provide free or affordable access to bicycles for short-distance trips in an urban area as an alternative to motorized vehicles, thereby reducing congestion, noise, and air pollution.
- Examples:
  - College park: mBike
  - Washington DC: Capital Bikeshare
  - NYC: Citi Bike – 12,000 bikes, 750 stations

Bike-sharing System



Bike-sharing System



Background on Motivate & Citi Bike

- Motivate operates systems in NYC, Chicago, Boston, Bay Area, Washington...
- Across systems, more than 100M rides since 2010
- Top 10 Most Innovative Companies in social good

Fast Company (2/18)

Citi Bike	
Stations	750+
Docks	25k
Bikes	12k
Subscribers	147k
Rides	17.5M

What is the greatest challenge?

Photo: Jim Henderson

Photo: “Another Believer”

Rebalancing Problem

Overnight Rebalancing

- Box trucks can hold about 50 bikes – sprinters about 20
- Rebalancing occurs in one shift 11pm – 6am
- Some stations not accessible to box trucks – use trikes

- Rebalancing costs alone may exceed a user‘s subscription fee
- Rebalancing cost: 10-15% of gross revenue

System Design

- Rebalancing is a major expenditure in bike-sharing
- Citi Bike would like to reallocate capacity to require less rebalancing
- How?

Citi Bike Data – Modeling Process

- Define Problem
  - Focus on rebalancing operations
  - Performance measure
  - Decision variables
- Data Analysis
  - Analyze the current operation – data exploration and statistical inference
  - Forecast demand – regression, time series
- Develop a Model for Decision Making
  - decision tree, optimization, simulation

Citi Bike Data: June 2017

1,048,576 observations.

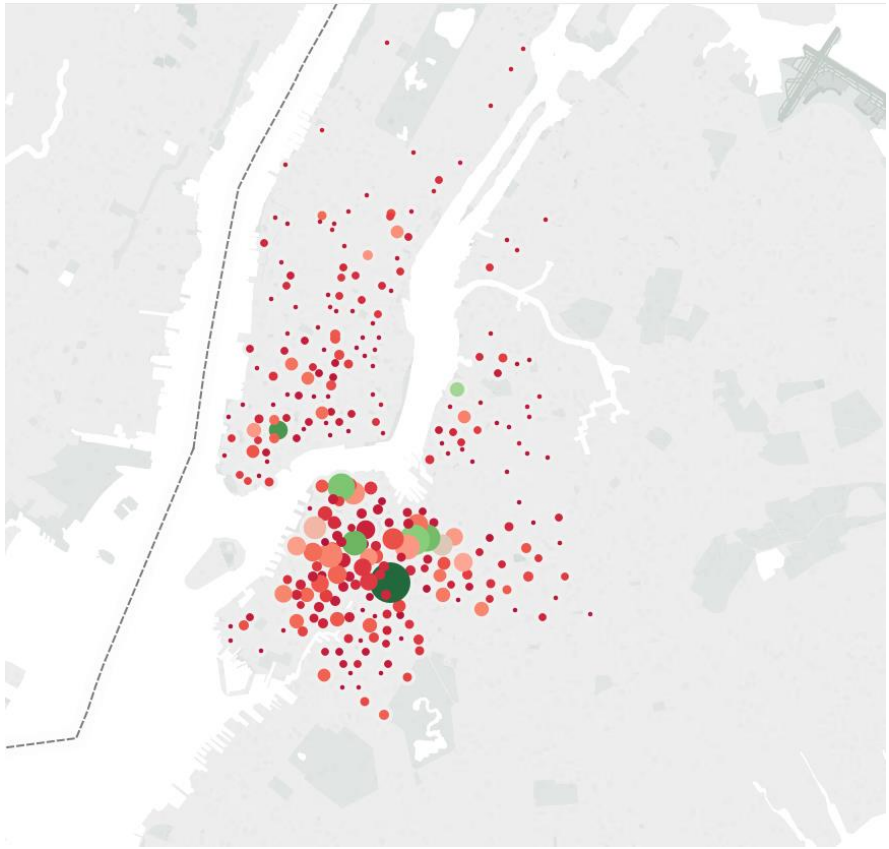
tripduration	starttime	stoptime	start station id	start station name	start station latitude	start station longitude
1397	6/1/2017 0:00	6/1/2017 0:23	515	W 43 St & 10 Ave	40.76009437	-73.99461843
1103	6/1/2017 0:00	6/1/2017 0:18	488	W 39 St & 9 Ave	40.75645824	-73.99372222
1810	6/1/2017 0:00	6/1/2017 0:30	461	E 20 St & 2 Ave	40.73587678	-73.98205027
1760	6/1/2017 0:00	6/1/2017 0:29	2009	Catherine St & Mor	40.71117444	-73.99682619
2165	6/1/2017 0:00	6/1/2017 0:36	360	William St & Pine S	40.70717936	-74.00887308
608	6					
416	6					
414	6					

end station id	end station name	end station latitude	end station longitude	bikeid	usertype	birth year	gender
3285	W 87 St & Amste	40.78839	-73.9747	26642	Subscriber	1967	1
297	E 15 St & 3 Ave	40.734232	-73.986923	25656	Subscriber	1981	1
465	Broadway & W 41	40.75513557	-73.98658032	21023	Subscriber	1982	1
527	E 33 St & 2 Ave	40.744023	-73.976056	25718	Subscriber	1973	2
474	5 Ave & E 29 St	40.7451677	-73.98683077	18691	Subscriber	1985	1
3386	1 Pl & Clinton St	40.6809591	-73.99905709	20729	Subscriber	1976	1
3255	8 Ave & W 31 St	40.75058535	-73.99468482	27900	Subscriber	1987	1
366	Clinton Ave & Myr	40.693261	-73.968896	20813	Subscriber	1985	1

Citi Bike Data Exploration

- On average, there were 57,705 rides per day, with each bike used 6.79 times per day.
- Visualize the data (using Tableau):

  Trip duration



Next …

- Descriptive data analysis (statistics)