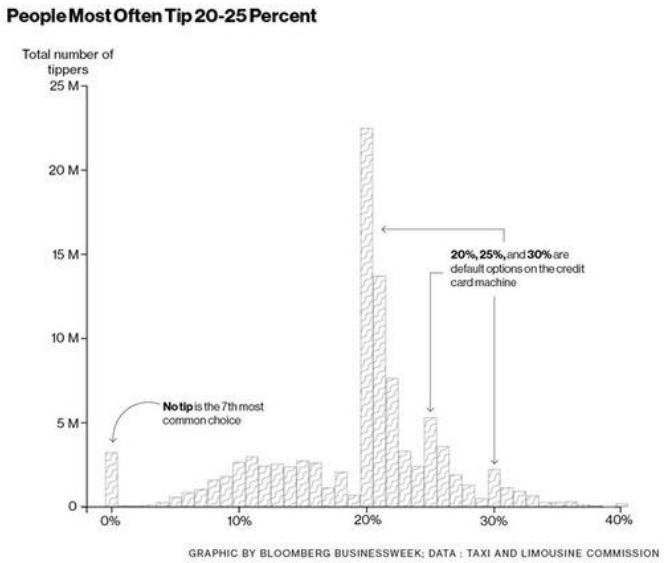
How Software in NYC Taxis Quietly Generates $5.2M in Extra Tips

A Critique of Data Journalism

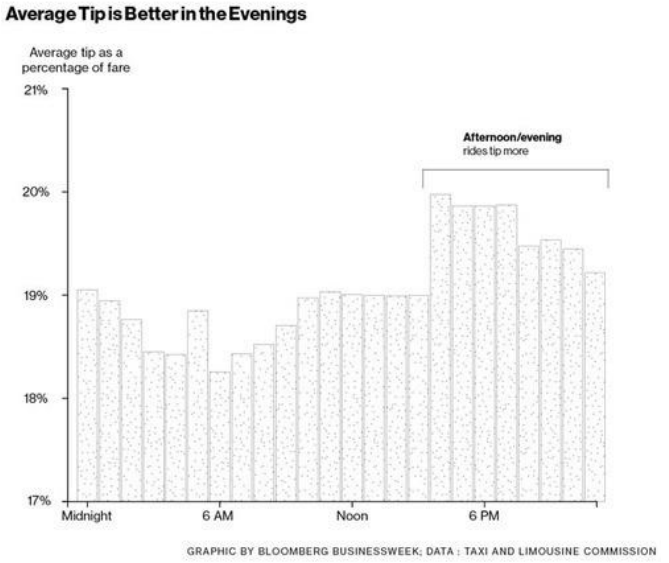
- Rahul Gandhi

Problem Statement

Businessweek article [2] (now on Yahoo Finance) claimed tipping spikes at 20-30% and a mysterious 4 p.m. surge. **But:** Their analysis ignored hidden variables (payment system vendors).



Businessweek's original Tip histogram (bumps at 20%, 25%, 30%).



Businessweek's original Average Tip Percentage Trends Throughout the Day

Data Sources

- **Primary data:** NYC Taxi and Limousine Commission (TLC) trip records from 2013, obtained via a Freedom of Information Law (FOIL) request by Chris Whong and shared publicly 2013. [3]
- **Supplemental data:** "I Quant NY" validated findings using personal cab receipts [1]

Data Processing

- Calculated tip percentages using two methods:
  - **Option 1**: Tip percentage based on *fare + surcharge*.
  - **Option 2**: Tip percentage based on *fare + surcharge + MTA tax + tolls*.
- Segmented data by payment system vendor (CMT vs. Verifone) to identify discrepancies.
- Re-analyzed aggregate statistics (e.g., average tip percentages by time of day) after correcting for vendor-specific calculations.

Data Validation

Examined individual ride records from the TLC dataset for both vendors separately (e.g., fare, surcharge, tolls, tip amount).

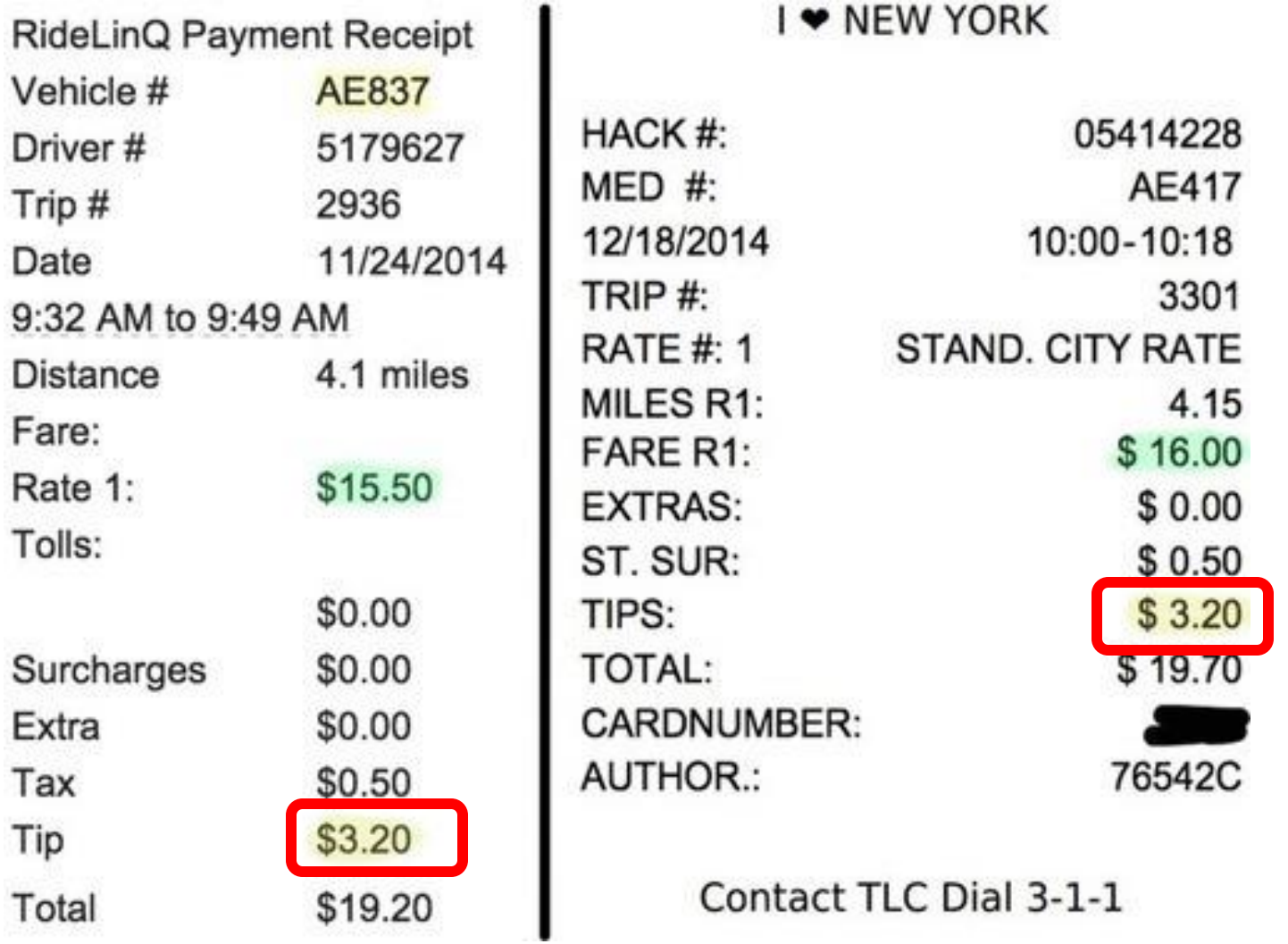
vendor	pickup	fare	srchrg	tax	tolls	total	tip	option 1	option 2
VTS	5/14/13 20:30	34.00	0.50	0.50	5.33	47.23	6.90	0.200	0.171
VTS	5/14/13 20:15	79.00	0.50	0.50	5.33	102.40	17.07	0.215	0.200
VTS	5/14/13 20:25	32.00	0.50	0.50	5.33	44.83	6.50	0.200	0.170
VTS	5/14/13 19:04	52.00	0.00	0.50	5.33	68.23	10.40	0.200	0.180
VTS	5/14/13 20:30	27.50	0.50	0.50	5.33	39.43	5.60	0.200	0.166
VTS	5/14/13 20:23	52.00	0.00	0.50	5.33	63.83	6.00	0.115	0.104
VTS	5/14/13 20:29	29.00	0.50	0.50	5.33	41.23	5.90	0.200	0.167
VTS	5/14/13 20:21	52.00	0.00	0.50	5.33	62.83	5.00	0.096	0.086
VTS	5/14/13 19:19	28.00	1.00	0.50	5.33	38.83	4.00	0.138	0.115
VTS	5/14/13 20:38	23.00	0.50	0.50	5.33	33.73	4.40	0.187	0.150
VTS	5/14/13 20:38	28.50	0.50	0.50	5.33	40.63	5.80	0.200	0.167
VTS	5/14/13 20:36	29.00	0.50	0.50	5.33	41.23	5.90	0.200	0.167
VTS	5/14/13 20:37	28.00	0.50	0.50	5.33	39.33	5.00	0.175	0.146
VTS	5/14/13 19:23	23.50	1.00	0.50	5.33	36.45	6.12	0.250	0.202
VTS	5/14/13 19:19	25.00	1.00	0.50	5.33	37.03	5.20	0.200	0.163
VTS	5/14/13 20:41	23.50	0.50	0.50	5.33	34.83	5.00	0.208	0.168
VTS	5/14/13 19:18	40.00	1.00	0.50	5.33	57.08	10.25	0.250	0.219
VTS	5/14/13 19:03	52.00	0.00	0.50	5.33	62.83	5.00	0.096	0.086
VTS	5/14/13 20:31	36.50	0.50	0.50	5.33	52.08	9.25	0.250	0.216
VTS	5/14/13 20:35	18.00	0.50	0.50	5.33	29.33	5.00	0.270	0.206
VTS	5/14/13 19:47	24.00	1.00	0.50	5.33	42.83	12.00	0.480	0.389

vendor	pickup	fare	srchrg	tax	tolls	total	tip	option 1	option 2
CMT	3/14/13 21:37	28.50	0.50	0.50	5.33	41.78	6.95	0.240	0.200
CMT	3/14/13 21:41	30.50	0.50	0.50	5.33	44.19	7.36	0.237	0.200
CMT	3/14/13 22:05	33.50	0.50	0.50	5.33	51.77	11.94	0.351	0.300
CMT	3/14/13 17:32	38.00	1.00	0.50	5.33	56.03	11.20	0.287	0.250
CMT	3/14/13 17:35	42.50	1.00	0.50	5.33	61.66	12.33	0.283	0.250
CMT	3/14/13 17:37	34.50	1.00	0.50	5.33	51.66	10.33	0.291	0.250
CMT	3/14/13 20:49	26.00	0.50	0.50	5.33	36.33	4.00	0.151	0.124
CMT	3/14/13 17:30	34.00	1.00	0.50	5.33	48.98	8.15	0.233	0.200
CMT	3/14/13 17:35	34.00	1.00	0.50	5.33	48.99	8.16	0.233	0.200
CMT	3/14/13 21:59	27.50	0.50	0.50	5.33	36.33	2.50	0.089	0.074

Data Validation

Cross-referenced receipts from personal taxi rides to confirm how payment systems (CMT - Left Side vs. Verifone - right side) calculated default tips.

For 20% Tip, Base fares in green are different. Here, CMT is calculating tips including tax, giving the driver the same default tip (**highlighted in red box**) for less mileage as the Verifone driver.



RideLinQ Payment Receipt

Vehicle #	AE837
Driver #	5179627
Trip #	2936
Date	11/24/2014
9:32 AM to 9:49 AM	
Distance	4.1 miles
Fare:	
Rate 1:	$15.50
Tolls:	
	$0.00
Surcharges	$0.00
Extra	$0.00
Tax	$0.50
Tip	$3.20
Total	$19.20

I ❤ NEW YORK

HACK #:	05414228
MED #:	AE417
12/18/2014	10:00-10:18
TRIP #:	3301
RATE #: 1	STAND. CITY RATE
MILES R1:	4.15
FARE R1:	$ 16.00
EXTRAS:	$ 0.00
ST. SUR:	$ 0.50
TIPS:	$ 3.20
TOTAL:	$ 19.70
CARDNUMBER:	[redacted]
AUTHOR.:	76542C

Contact TLC Dial 3-1-1

Tools used

- No specific analysis tools mentioned.
- **Likely tools include SQL/Python/R** for data manipulation
- **Spreadsheets** for sample ride analysis (Based on Image attached)
- **Likely visualization tools** (e.g., matplotlib, Tableau) for histograms

Key Findings

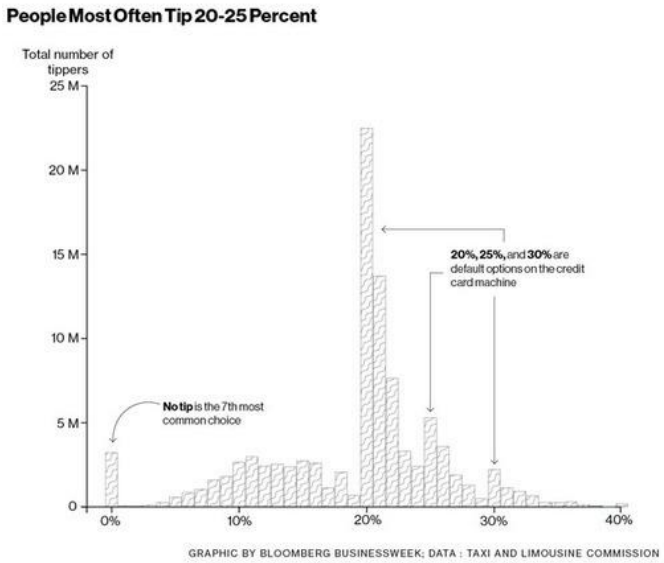
- **Vendor Discrepancy:** CMT systems included taxes/tolls in tips; Verifone did not [1].
- **$5.2M/year Extra Tips:** From CMT riders using defaults tip options (20%, 25% and 30%).
- **Additional Tip Amount:** ~$200 per year for each CMT Driver compared to Verifone
- **Flawed original analysis**: *Businessweek*'s conclusion about rush-hour tipping behaviour ignored vendor differences and surcharge timing
- **4 p.m. "Surge" Debunked:** Increased in % Tip caused by rush-hour surcharges, not rider empathy [1][2]

Takeaways

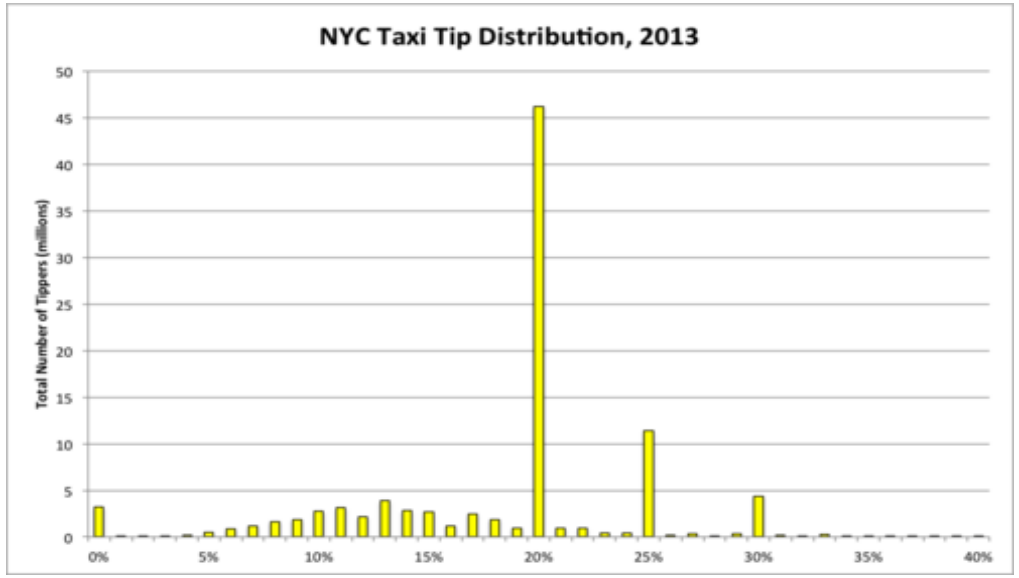
- Payment system inconsistencies disadvantage riders and drivers.
- Raw data inspection is essential to uncover hidden variables.

Visual Comparison

Original vs. Corrected Tip distributions



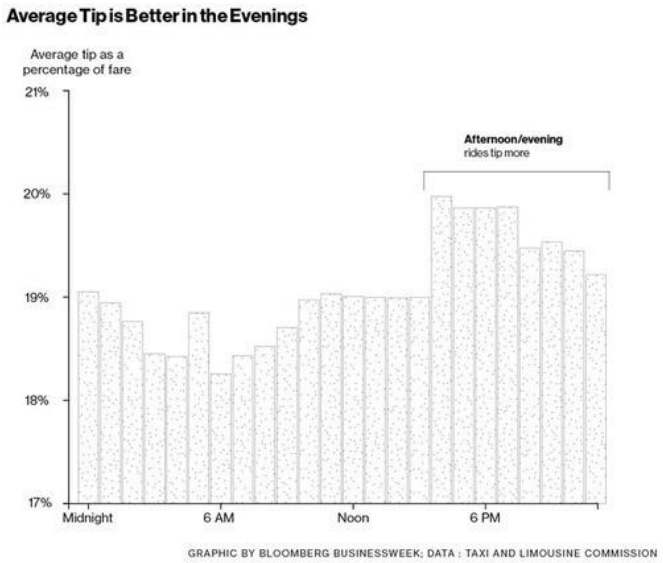
Businessweek's original Tip histogram (bumps at 20%, 25%, 30%).



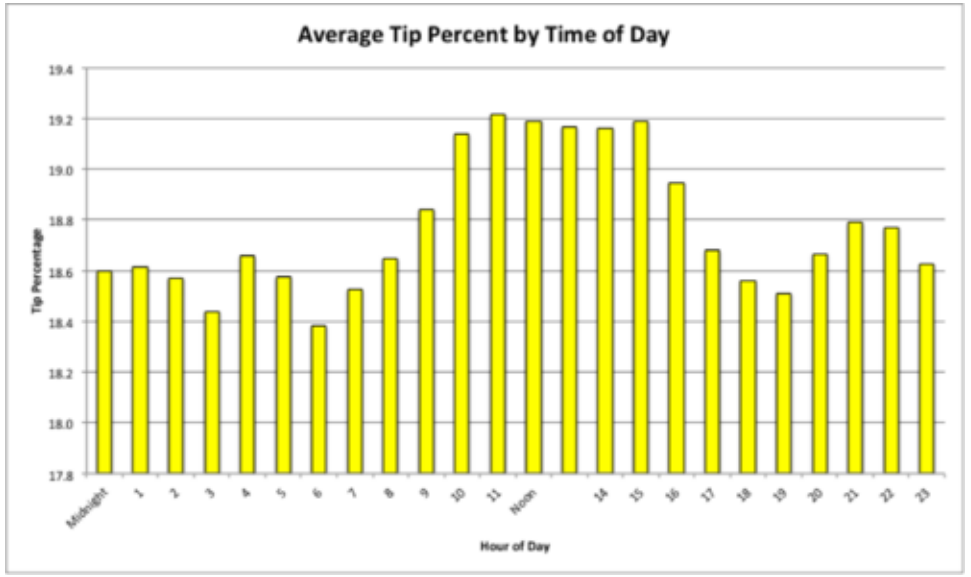
I Quant NY"'s corrected histogram (vendor split)

Visual Comparison

Original vs. Corrected Average Tip Percentage Trends Throughout the Day



Businessweek's original Tip Percentage Trends Throughout the Day



I Quant NY"'s corrected Tip Percentage Trends Throughout the Day

Strengths of Analysis and Visualizations

Granular Data Examination: Analyzed individual ride receipts to uncover vendor-specific tip logic (CMT vs. Verifone).

Transparency: Publicly shared methodology and leveraged open data for reproducibility.

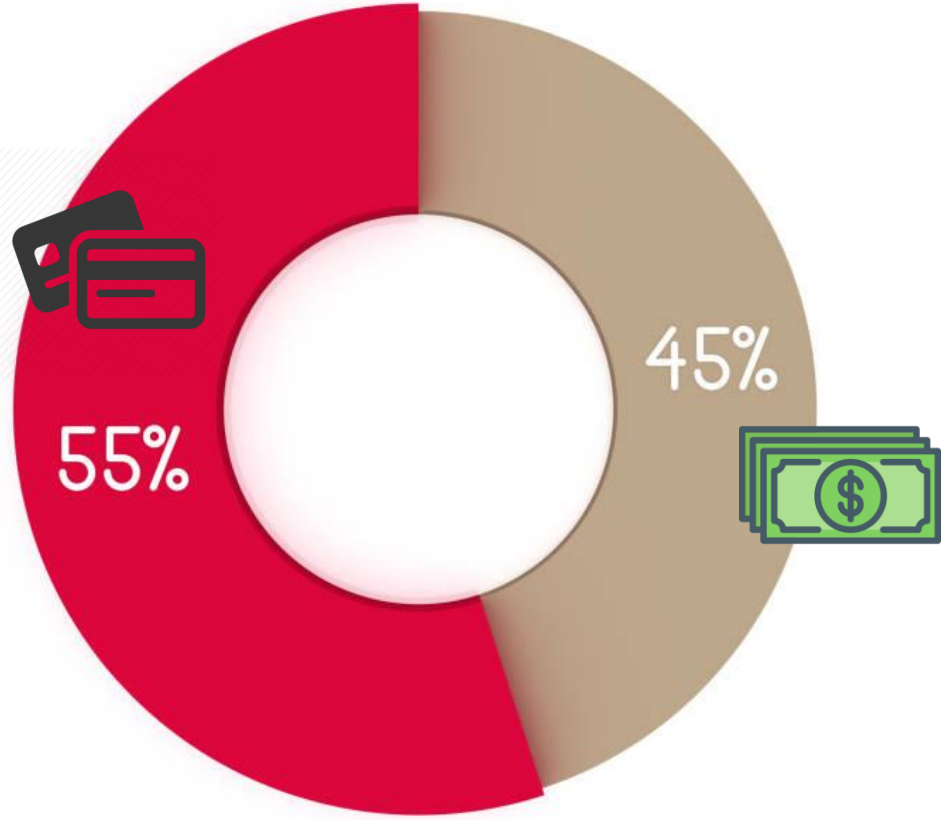
Impactful Visualizations: Corrected histograms (vendor-split tip percentages) resolved anomalies in the original analysis.

Policy Advocacy: Highlighted the need for TLC to standardize payment systems and release data openly

Weakness of Analysis and Visualizations

Excluded Cash Transactions: Analyzed 55% credit card transactions only, ignoring 45% cash tips [2]

- **Risk:** Overgeneralizes findings (e.g., default button impacts) to all riders, inflating credit card-specific quirks
- **Gap:** Fails to explore whether cash riders face similar vendor inconsistencies (e.g., fare rounding).



Weakness of Analysis and Visualizations

No Temporal Sensitivity Analysis

- **Static Timeframe**: Analyzed only **2013 data**, ignoring trends over time to understand shifts in tipping habits or payment tech updates.
- **Seasonal Blindspot**: Did not test for seasonal variations (e.g., holiday generosity vs. summer tourism).

Lack of Policy Impact Assessment: Did not publish follow-up article on the impact of this advocacy on real-world policy changes.

Thank You

References

[1] I Quant NY, "How Software in Half of NYC Cabs Generates $5.2 Million a Year in Extra Tips," *Tumblr*, 2015. [Online]. Available: https://iquantny.tumblr.com/post/107245431809/how-software-in-half-of-nyc-cabs-generates-52 [Accessed: Feb. 1, 2025].

[2] "Here's How Much You Should Be Tipping Your Cab Driver," Yahoo Finance, 2014. [Online]. Available: https://finance.yahoo.com/news/heres-much-tipping-cab-driver-183641123.html. [Accessed: Feb. 4, 2025]

[3] C. Whong, "FOILing NYC's Taxi Trip Data," Mar. 18, 2014. [Online]. Available: http://chriswhong.com/open-data/foil_nyc_taxi/. [Accessed: Feb. 4, 2025].