

The background is a blurred image of a financial market data screen. It features various stock indices and their values in different colors (green for up, red for down). Visible text includes "OMXRG1", "OMX ICELAND 8", "27956.04", "1632.51", "6230.9", and "1172.94". A white rounded rectangle is superimposed over the center of the image, containing the title and authors.

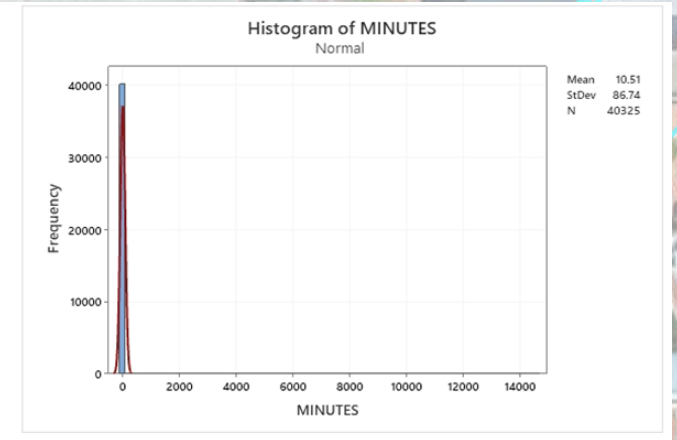
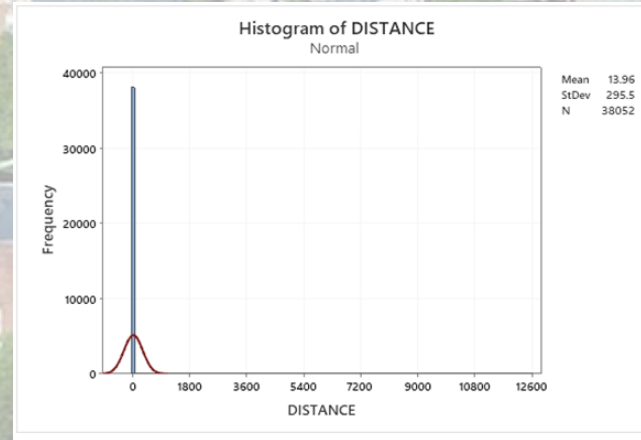
UMD Data Challenge 2021

Ashish Bachavala & Matthew Braun

DATASET

- More than 40,000 entries
- Unique Ride_ID Primary Key
- Mainly Floating-Point data types
- Filled with errors
- Repetitive Data
- Large range of values for Distance, Time, and number of coordinates

Before

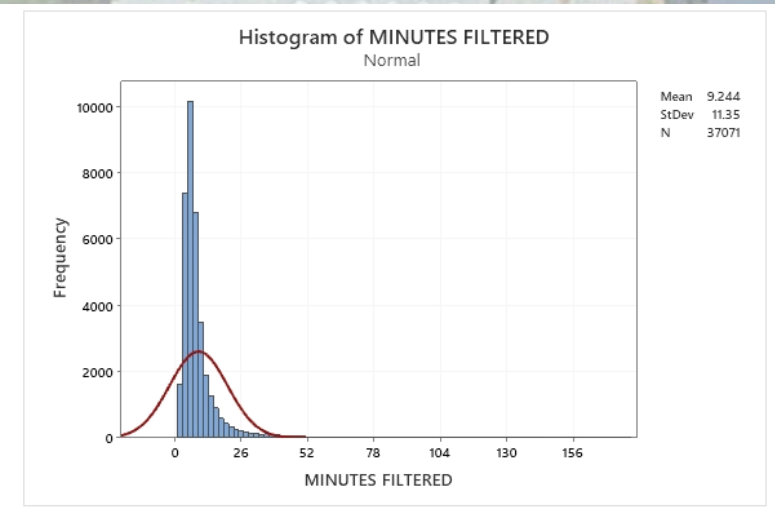
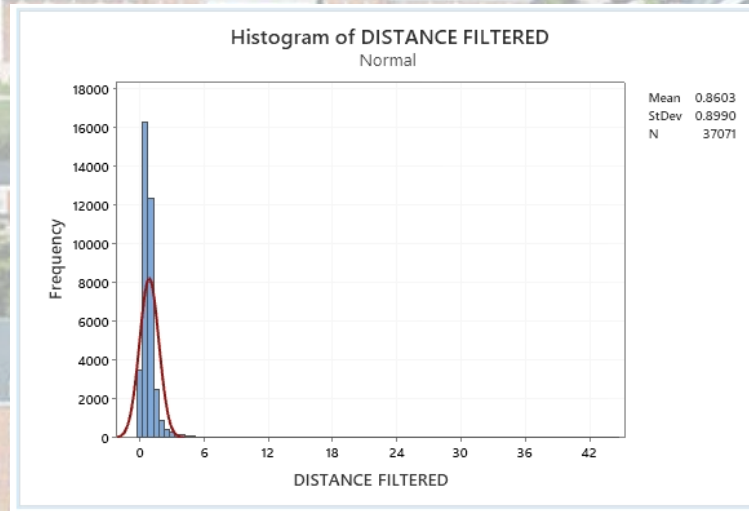


veo

WHAT WE DID

- Removed erroneous data through criteria requirements
- Calculated which points are in or out of campus

After

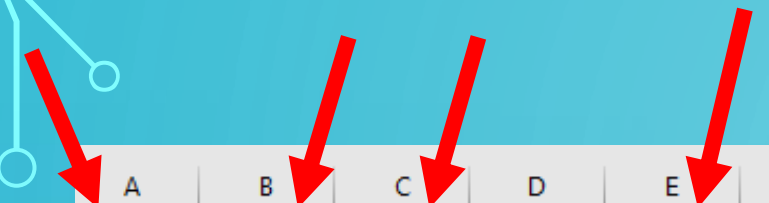


```
else:
    path = json.loads(row[10])
    path_len = len(path["coordinates"])
    print("Index: {} \t Distance: {} \t Minutes: {} \t Length of path: {}".format(row[0], row[5], row[8], path_len), end="\t")
    keep = path_len > 1 and int(row[8]) < 180 and int(row[8]) > 0 and float(row[5]) < 50
    if keep:
```

```
else:
    start_inside = point_in_polygon(boundaries, [float(row[12]), float(row[13])])
    end_inside = point_in_polygon(boundaries, [float(row[14]), float(row[15])])
```


veo

Before



A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
	Unnamed	CREATED	START	END	DISTANCE	RIDE_ID	CUSTOMER_ID	MINUTES	VEHICLE_TYPE	PATH	TIMESTAMP	START LONG	START LAT	END LONG	END LAT
0	0	2020-10-03	2020-10-03 03:32:15+00:00	2020-10-03 03:07:39+00:00	0.008	3135665	193964	1	Scooter	{	[-76.9342	38.99066	-76.9343	38.99074
1	1	2020-10-03	2020-10-03 03:07:39+00:00	2020-10-03 03:17:05+00:00	1.062	3135395	542947	30	Scooter	{	[-76.9382	38.98209	-76.9492	38.99285
2	2	2020-10-03	2020-10-03 03:17:05+00:00	2020-10-03 02:24:58+00:00	0.847	3135503	517321	8	Scooter	{	[-76.9376	38.97961	-76.9333	38.99268
3	3	2020-10-03	2020-10-03 02:24:58+00:00	2020-10-03 03:19:34+00:00	3.447	3134955	578051	47	Scooter	{	[-76.9436	38.98247	-76.9435	38.98248
4	4	2020-10-03	2020-10-03 03:19:34+00:00		0.386	3135529	552305	4	Scooter	{	[-76.9381	38.98208	-76.9433	38.9861

After



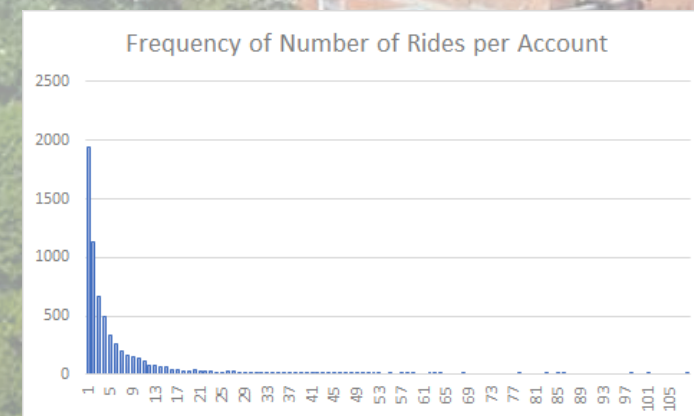
A	B	C	D	E	F	G	H	I	J	K	L	M	N
YEAR	START	DISTANCE	RIDE_ID	CUSTOMER_ID	MINUTES	VEHICLE_TYPE	START LONG	START LAT	START INSIDE	END LONG	END LAT	END INSIDE	MONEY
2020	2020-10-03 03:32:15+00:00	0.008	3135665	193964	1	Scooter	-76.934163	38.9906555	FALSE	-76.93428483	38.99073867	FALSE	\$ 1.25
2020	2020-10-03 03:07:39+00:00	1.062	3135395	542947	30	Scooter	-76.93818083	38.98208717	TRUE	-76.94916483	38.99284867	TRUE	\$ 8.50
2020	2020-10-03 03:17:05+00:00	0.847	3135503	517321	8	Scooter	-76.93759983	38.97961267	FALSE	-76.93333467	38.992676	FALSE	\$ 3.00
2020	2020-10-03 02:24:58+00:00	3.447	3134955	578051	47	Scooter	-76.94355317	38.98247017	TRUE	-76.94354633	38.98247567	TRUE	\$12.75
2020	2020-10-03 03:19:34+00:00	0.386	3135529	552305	4	Scooter	-76.93811567	38.9820835	TRUE	-76.943257	38.986101	TRUE	\$ 2.00

FINDINGS

- Average ride is 0.86 miles in 9.24 minutes and generates \$3.29
- More than 40% of rides start and end on campus

Row Labels	Average of DISTANCE	Average of MINUTES	Average of MONEY
2019	0.876741313	10.04907211	3.466284718
2020	0.847706229	8.628943987	3.14634274
Grand Total	0.860283321	9.244099161	3.284931618

Count of RIDE_ID	Column Labels		Start Outside Total		Start Inside		Start Inside Total
Row Labels	End Outside	End Inside	End Outside	End Inside	End Outside	End Inside	
2019	18.61%	20.36%	38.97%	20.01%	41.01%	61.03%	
2020	19.87%	17.73%	37.60%	17.87%	44.53%	62.40%	
Grand Total	19.33%	18.87%	38.19%	18.80%	43.01%	61.81%	



REVENUE ANALYSIS

Row Labels ▾	Average of MONEY	Average of MINUTES	Average of DISTANCE
E-Bike	\$ 3.15	10.77	0.91
Scooter	\$ 3.29	9.17	0.86
Grand Total	3.284931618	9.244099161	0.860283321

Row Labels ▾	Average of MONEY	Rides	Sum of MONEY
379235	\$2.81	108	\$303.75
575343	\$2.52	101	\$255.00
327797	\$2.76	101	\$279.00
296373	\$2.81	98	\$275.75
529959	\$2.36	86	\$203.00
393922	\$3.13	85	\$266.00
542926	\$2.46	83	\$203.95
423321	\$2.56	78	\$200.00

START INSIDE	FALSE	▾
END INSIDE	FALSE	▾
Row Labels ▾	Average of MONEY	
2019	\$4.12	
2020	\$3.78	
Grand Total	\$3.92	

START INSIDE	TRUE	▾
END INSIDE	TRUE	▾
Row Labels ▾	Average of MONEY	
2019	\$3.26	
2020	\$2.89	
Grand Total	\$3.04	

Sum of MONEY	Column Labels ▾	
Row Labels ▾	End Outside	End Inside
2019	41.70%	58.30%
2020	41.30%	58.70%
Grand Total	41.49%	58.51%

Sum of MONEY	Column Labels ▾	
Row Labels ▾	Start Outside	Start Inside
2019	41.84%	58.16%
2020	41.70%	58.30%
Grand Total	41.77%	58.23%

MAPS

DC 2021 - TEAM 10 (ASHISHBACH.COM)

INTERESTING INSIGHTS/PATTERNS

- From 2019 to 2020: On campus rides increased in distance, but off campus rides decreased

Sum of MONEY Column Labels ▾		
Row Labels ▾	E-Bike	Scooter
2019	8.03%	91.97%
2020	1.81%	98.19%
Grand Total	4.65%	95.35%

Count of VEHICLE_TYPE Column Labels ▾		
Row Labels ▾	E-Bike	Scooter
2019	9.44%	90.56%
2020	1.33%	98.67%
Grand Total	4.84%	95.16%

START INSIDE	TRUE	⌵
END INSIDE	FALSE	⌵
Row Labels ▾	Average of MINUTES	
2019	9.732731798	
2020	8.301464714	
Grand Total	8.96154398	

PROBLEMS WE RAN INTO



Limitations of single-thread Python



Scraping of campus coordinates



THANK YOU FOR LISTENING

Questions?

Programs/Languages used:

- Python
- Excel
- Minitab
- HTML

veo