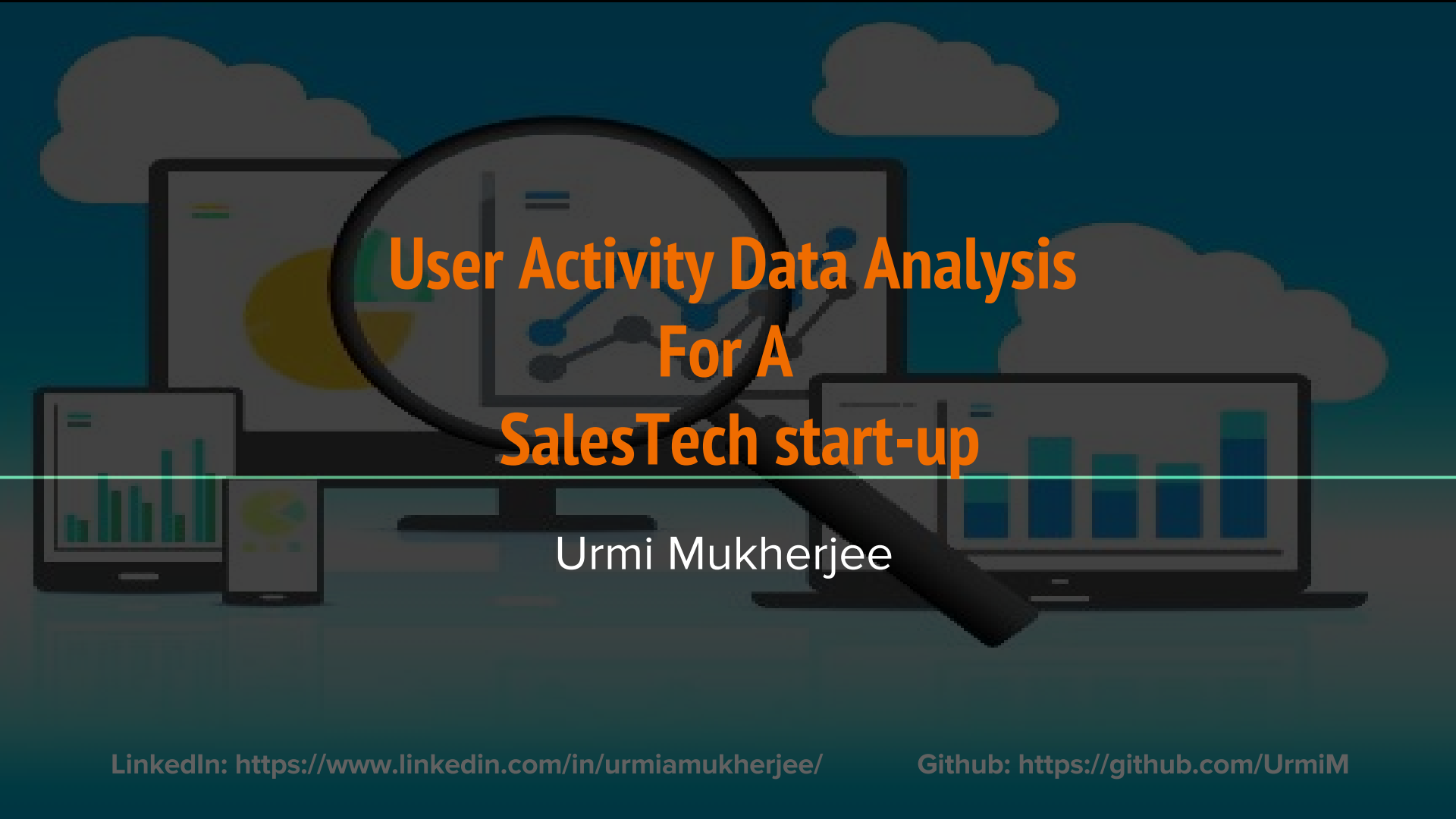


User Activity Data Analysis For A SalesTech start-up

Urmi Mukherjee

LinkedIn: https://www.linkedin.com/in/urmiamukherjee/ Github: https://github.com/UrmiM

Company Introduction

- SalesTech Start Up for Financial Services
- Empowers companies to engage clients and sell financial products over Text Messaging
- Allows client-facing teams to leverage Messaging platforms like iMessage, SMS, Facebook Messenger etc. to engage prospects and clients

What can the data tell us?

- Behavior of the customer differ based on geographical locations, industry type, time of day, day of week, month?
- Identify on-the-fence customers? How soon?
- Underlying "View" patterns specific to a certain industry type?
- Generalize “View” patterns for all customers?
- Help the business leaders in focusing and prioritizing efforts and brain power?
- Cluster all the users based on viewing device?
- Relationship between view duration, number of views, viewing device & browser?

Data Format

Activity Type	User ID	RM ID	User OS	Viewing App	User Groups	Shared Date	Shared Time	Title of Content	Share Channel	Viewed	Time Spent in Seconds	Last Activity Date	Last Activity Time
SHARE	5931852	197521			Client Solutions	18-Aug-17	0:53:39	English/Writing GameScapes and Games	WEB	1	5505	18-Aug-17	22:43:00
VIEW	5931852		WINDOWS_7	CHROME				Introductions and Conclusions: Set the Stage		1	285	18-Aug-17	22:37:45
VIEW	5931852		WINDOWS_7	CHROME				Paragraph Structure: Find the Balance		1	1635	18-Aug-17	22:00:03
VIEW	5931852		WINDOWS_7	CHROME				Paragraph Structure: Find the Balance		1	3555	18-Aug-17	21:00:01
VIEW	5931852		WINDOWS_7	CHROME				Paragraph Structure: Find the Balance		1	30	18-Aug-17	20:59:28
SHARE	5864595	197521			Client Solutions	10-Aug-17	3:27:06	English/Writing GameScapes and Games	WEB	1	1240	18-Aug-17	20:56:00
VIEW	5864595		WINDOWS_7	IE11				Thesis Statements: What's the Point?		1	5	18-Aug-17	20:55:11
SHARE	5926033	197521			Client Solutions	17-Aug-17	21:41:24	Business Foundations	WEB	0	0	17-Aug-17	21:41:24
SHARE	3837716	197521			Client Solutions	17-Aug-17	21:33:05	Business Foundations	WEB	1	51490	17-Aug-17	21:33:05
VIEW	197521		WINDOWS_10	CHROME				DEMO - Planning and Writing		1	350	17-Aug-17	3:11:48
VIEW	197521		WINDOWS_10	CHROME				Topics and Learning Objectives		1	35	17-Aug-17	3:08:54
VIEW	197521		WINDOWS_10	CHROME				Topics and Learning Objectives		1	25	17-Aug-17	3:04:10

- **Data of 4 customers from different industries**
- **Each row is an action. Share/View**
- **All Categorical features except “Time Spent in Seconds”**

Activity Spread Across Customers

Customer	IS	ES	TW	YB
No. of Users (Share)	434	235	263	100
Avg. Shares/User	2.54	1.89	1.44	1.40
Avg. Views/Share	2.86	15.26	9.28	3.06

-- **Avg. Shares/User measures of customer engagement with the platform**

- **Customer IS shows maximum engagement.**

-- **Avg. Views/Share is the clients' interest with the content shared**

- **This completely depends on the file type, size and content**

Customer Engagement Metrics

Customer-->	IS		ES		TW		YB	
Month	No. of Users	No.of Shares/User	No. of Users	No.of Shares/User	No. of Users	No.of Shares/User	No. of Users	No.of Shares/User
Feb	N/A	N/A	18	1.22	55	1.00	N/A	N/A
Mar	96	1.85	57	1.37	92	1.10	N/A	N/A
Apr	70	2.29	49	1.47	59	1.68	N/A	N/A
May	92	2.01	51	1.61	25	1.32	N/A	N/A
Jun	91	1.99	48	1.50	27	1.26	60	1.48
Jul	77	1.82	52	1.46	23	1.74	12	1.17
Aug	101	2.57	33	1.24	15	1.07	31	1.19

Predicting “View_time_spent” to identify On-The-Fence clients

- **Predict “Time spent” using Linear Regression, very low r-squared value 0.06, very low signal**
- **Used Random-Forest-Regressor for the same feature set, no improvement in the r-squared value**
- **Fitted a Logistic Classification Model, by assuming 20% randomly selected users with number of views greater than the mean converted, accuracy of about 85%**
- **New feature which combines number of views and type of document viewed would give more signal**

About the Data

- The data has 6000 rows but total number of sharing-users is 86 and viewing-users is 600
- Predicting “Time Spent In Views” not possible since no information about content viewed
- Classification not possible
- Why fit a model on hypothetical data?

Recommendations

- **Collect user conversion data**
- **Content is very important to predict views and identify on-the-fence customers**
- **File type, size and content information should be collected**
- **User specific data to analyse and find user patterns**

I’d like to Thank --

-- Eltropy
-- Scrum Leaders
-- Instructors & DSRs
-- Cohort Mates
-- Galvanize Inc

Thank You!

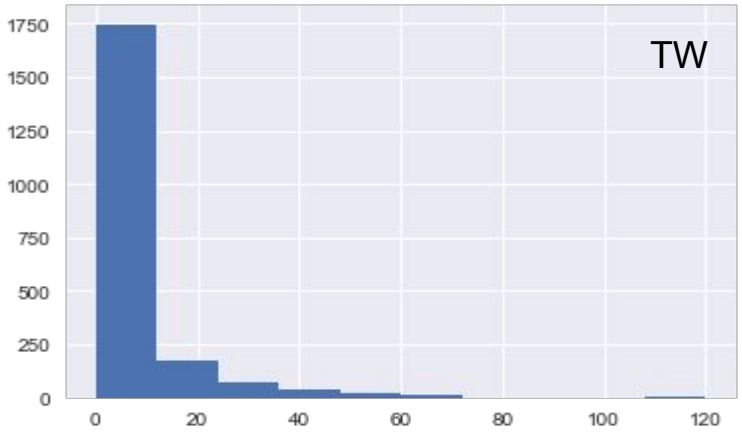
Questions?

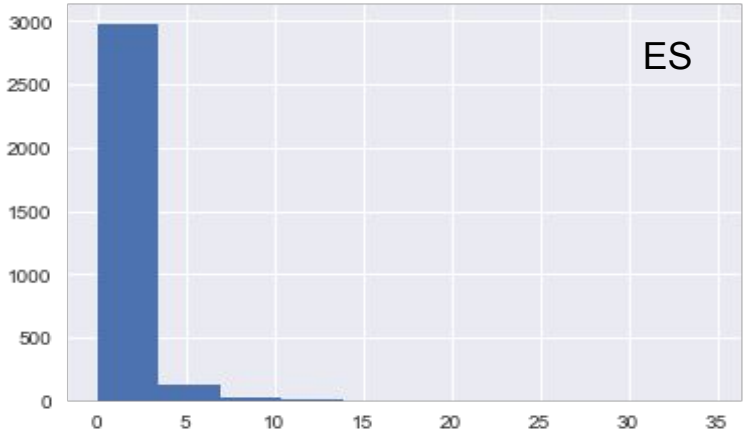
Email: Urmiamukherjee@gmail.com

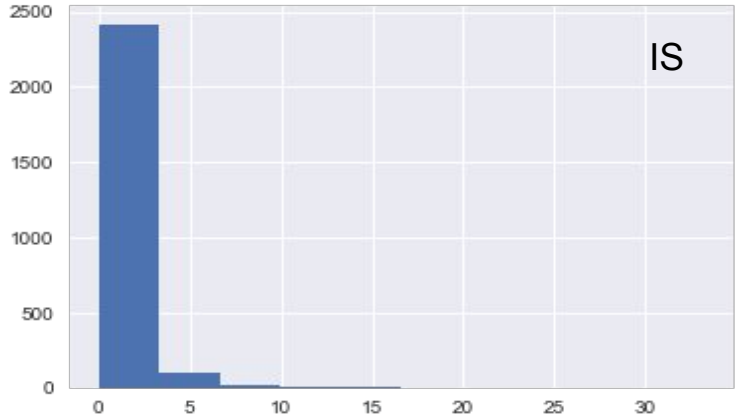
LinkedIn: https://www.linkedin.com/in/urmiamukherjee/

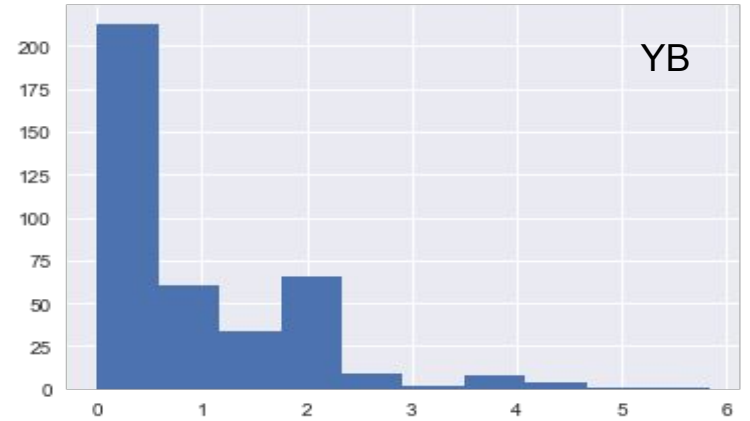
Github: https://github.com/UrmiM

Appendix I - Total View Time Spent across Companies
X-Axis - Time in minutes, Y-Axis - No of views

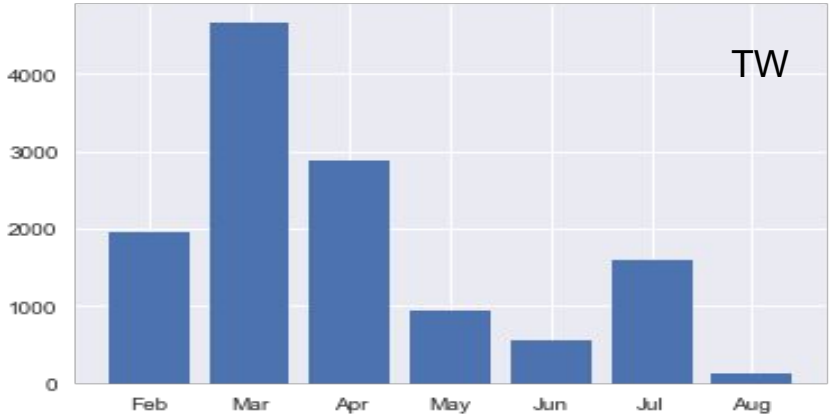


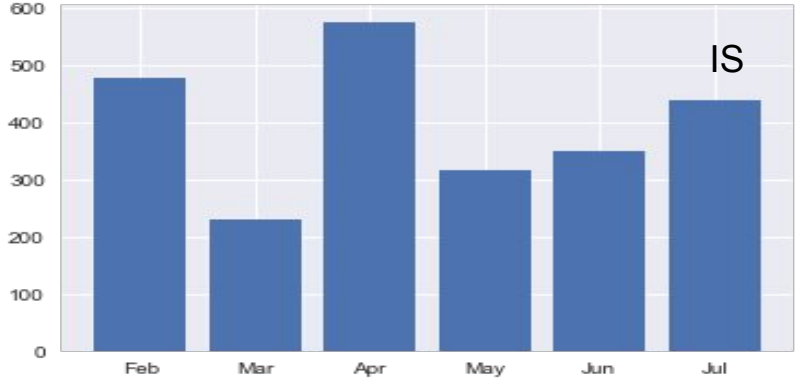


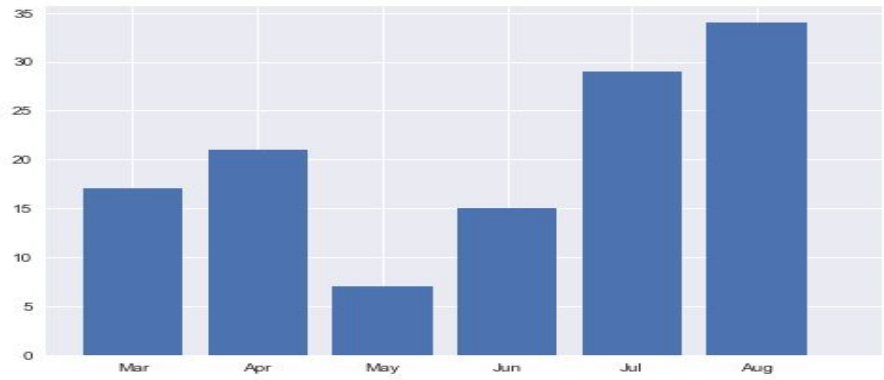




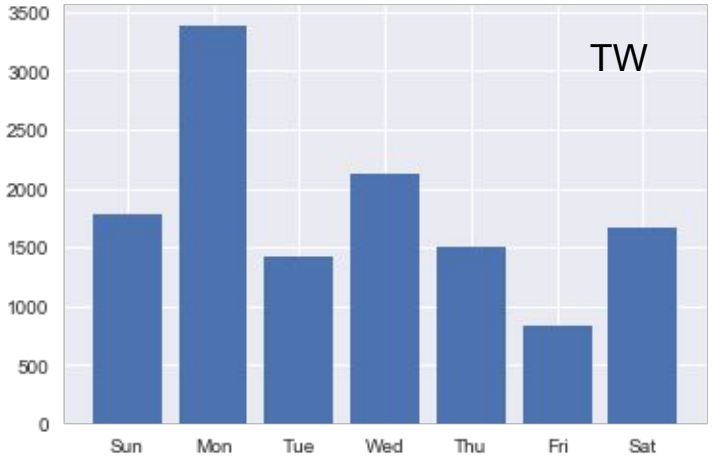
Appendix II - Total View Time Spent across Companies per Month
X-Axis - Months, Y-Axis - Total minutes spent in views

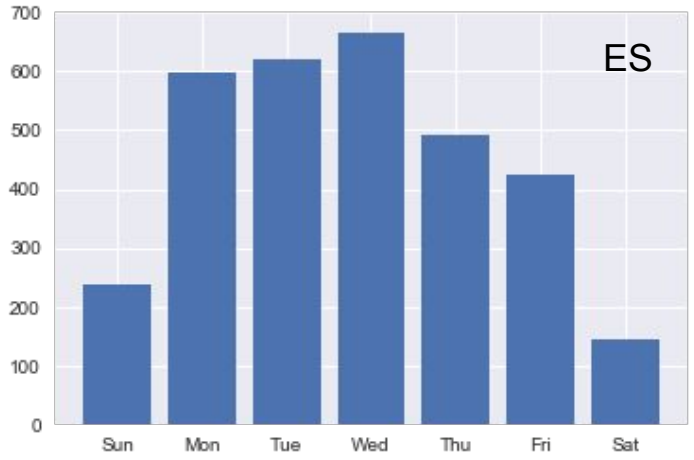


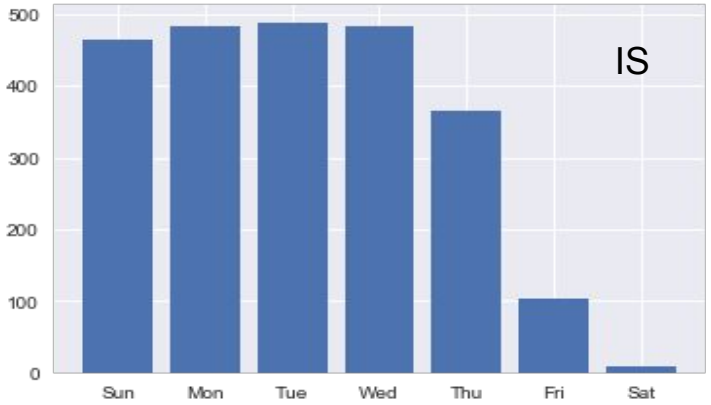


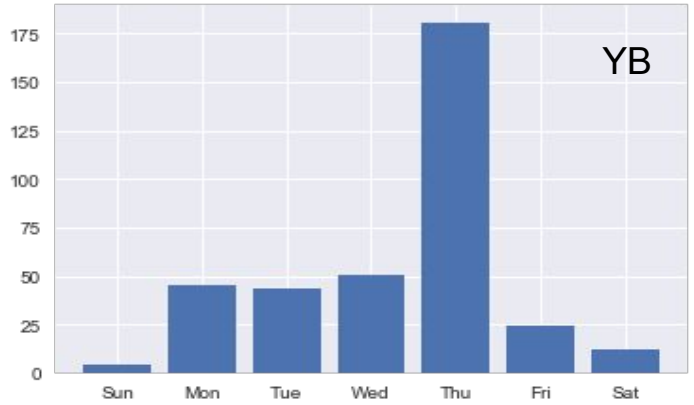


Appendix III - Total View Time Spent across Companies per Day-of-week
X-Axis - Day-of-week, Y-Axis - Total minutes spent in views



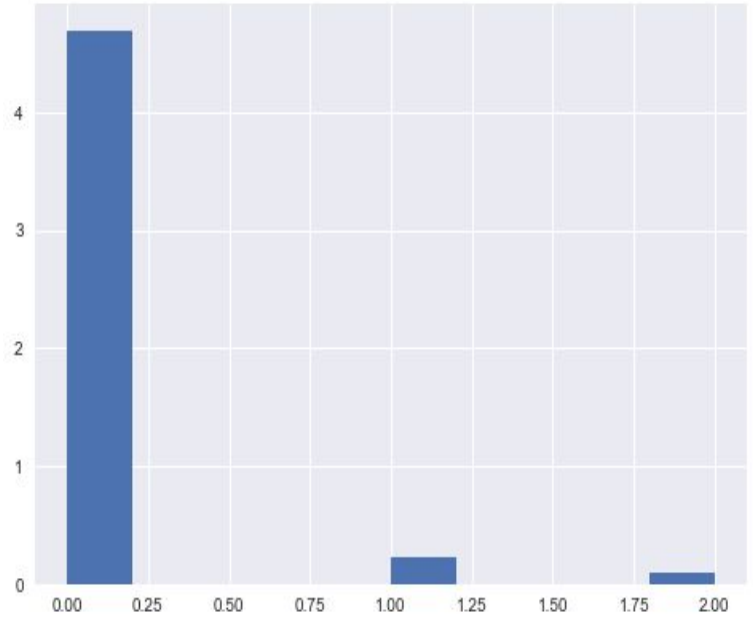






Customer YB - Behavior Anomaly

View Time



Day of Week