MODULE 3 FINAL PROJECT

BY ÜLKÜ & KIRYL

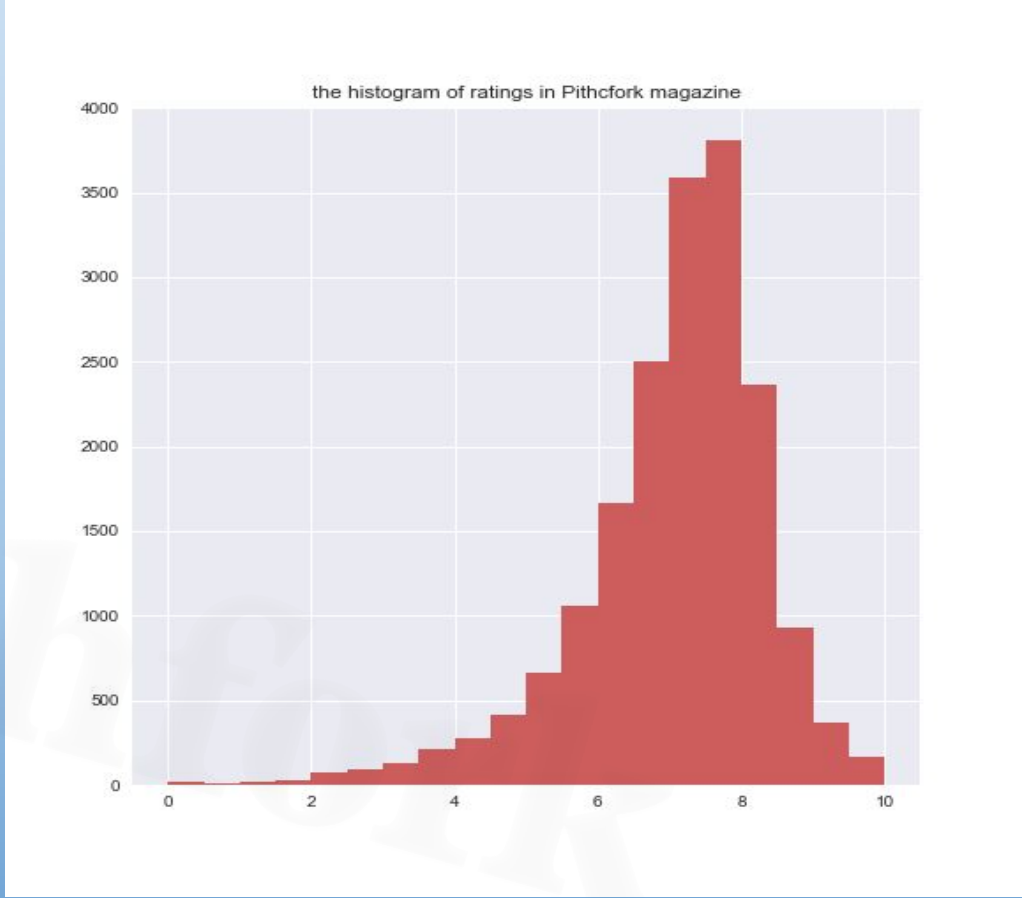
06/11/2019

DataSet Characteristics

'Pitchfork' is a music-centric online magazine. It was started in 1995 and grew out of independent music reviewing into a general publication format, but is still famed for its variety music reviews.

	score
count	18393.000000
mean	7.005779
std	1.293675
min	0.000000
25%	6.400000
50%	7.200000
75%	7.800000
max	10.000000

Data Points of Total Ratings



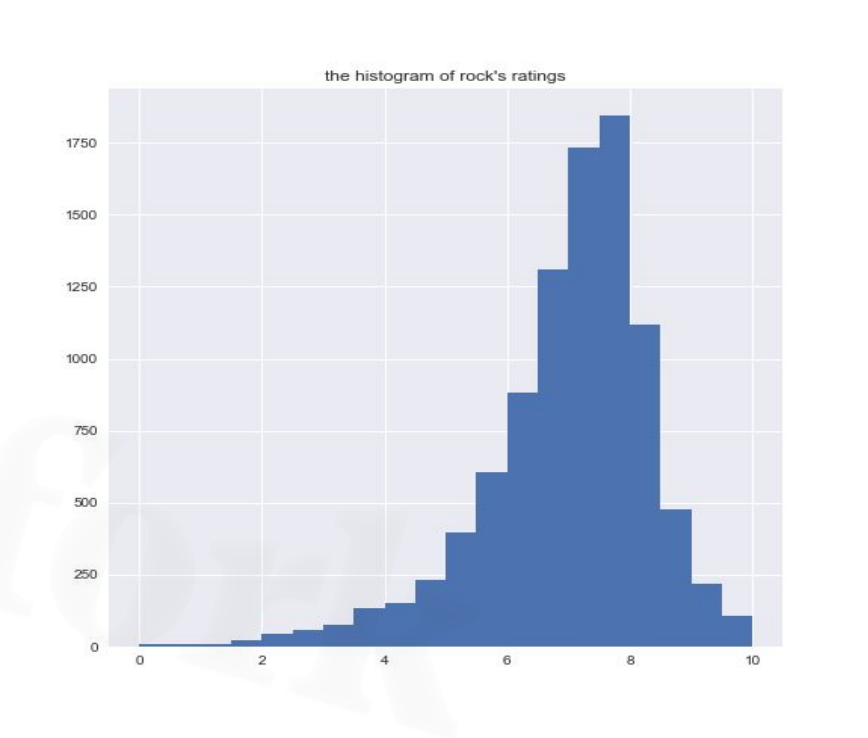
***We assume we work with continuous and independent data set*

Test #1: One Sample T-Test (left side)

Question: *Is there a difference between the ratings of ROCK music and all other music genres?*

Goal: Test, at 5% level of significance, whether are there differences in magazine's ratings with respect to genres. In particular test, is rock genre treated the same as other music genres.

Hypothesis:
H0= There is **no significant difference** in ratings of rock music comparing to others.
HA= The ratings of rock music **are** usually **lower** than ratings of all genres.



T-crit = - 1.645 at alpha = 0.05

T statistics = -2.815

Conclusion:

Because t-statistics is less than t-critical, i.e. is located to the left of t-critical value with alpha = 0.05, we reject the null hypothesis and can conclude that rock music in general rated lower.

Test #2: Two Sample T-Test

***Question**: Is there a statistical difference between the average ratings of jazz music and metal music genres?*

	Jazz Music Score		Metal Music Score
count	435.000000	count	862.000000
mean	7.303908	mean	6.950000
std	1.251104	std	1.387799
min	1.000000	min	0.200000
25%	6.900000	25%	6.500000
50%	7.600000	50%	7.300000
75%	8.000000	75%	7.900000
max	10.000000	max	9.700000

Goal:

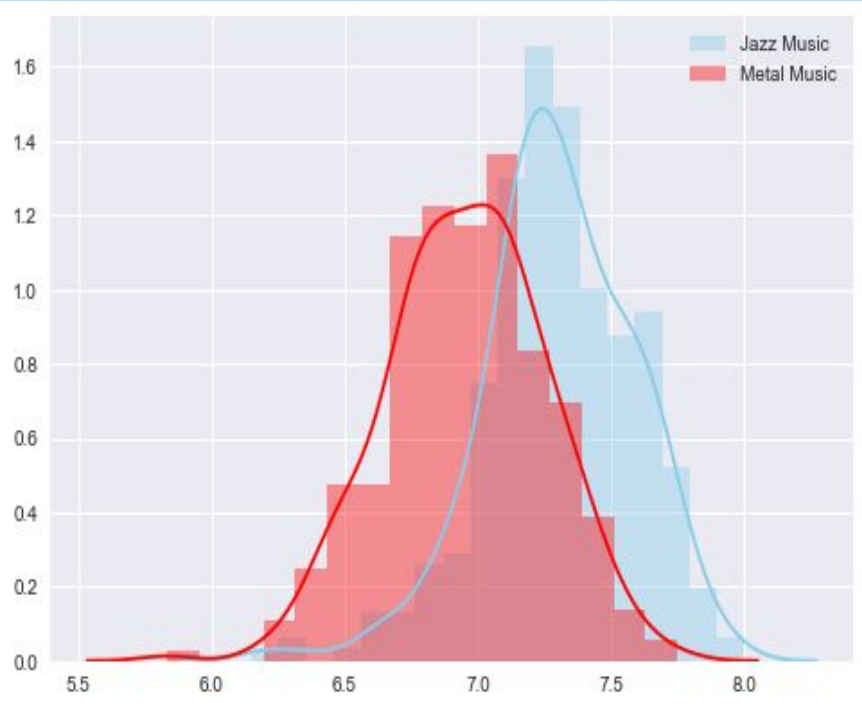
Test, at 5% level of significance, whether are there differences in average ratings of two different music genres. In particular test, is jazz music and metal music.

Hypothesis:

H0: There is no significant difference between mean values of jazz and metal data set.

HA: There is a significant difference between jazz and metal mean values.

Results of Two Sample T-Test: T-statistics: -14.57 and P-value: 0.0



T-critical = - 1.65 at alpha = 0.05

T statistics = -14.57

Conclusion:

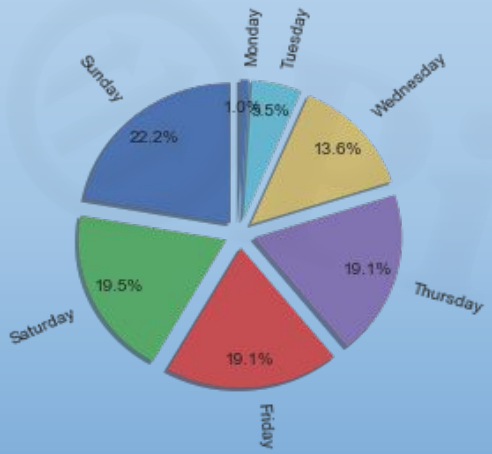
Because t-statistics is less than t-critical, i.e. is located to the left of t-critical value with alpha = 0.05 level of significance, we reject the null hypothesis and can conclude that jazz and metal music average ratings are statistically different.

Test # 3: One-Factor ANOVA Test

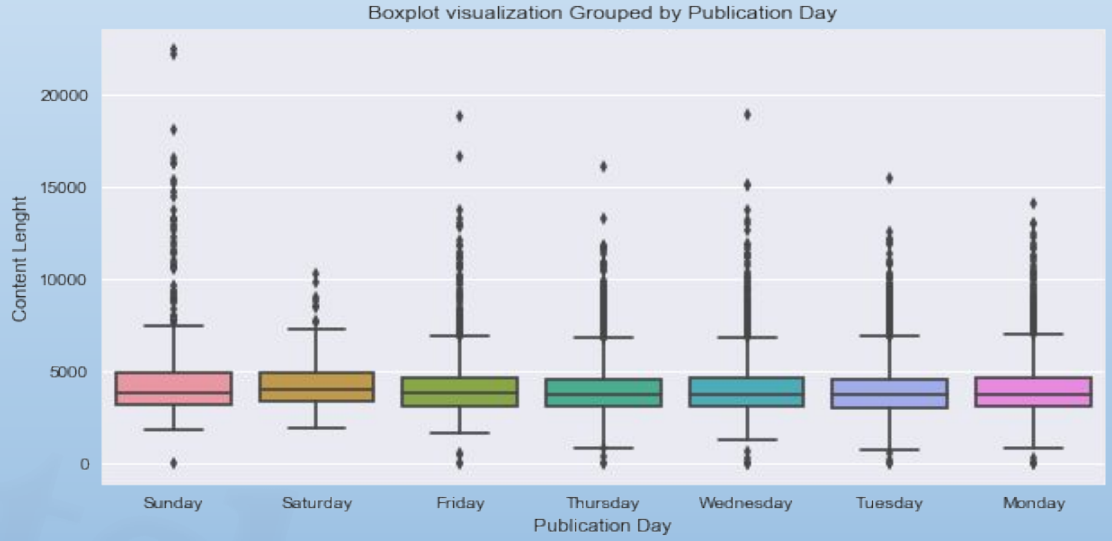
***Question**: Is there a significant effect of publication day on content length?*

Goal:

Test, at 5 % level of significance, whether there is a significant effect of review publication day on review content length?



Proportions of Content Publications by Day



Hypothesis:

H0: There **is no** significant **difference** in average content length depending on which day it is published.

HA: There **is** a significant **difference** in average content length depending on which day it is published.

At 5% level of confidence with 6.0 degrees of freedom, SSM for the model 1.429448e+08, F= 11.67 and $p < 0.05$.

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	4094.9888	28.543	143.468	0.000	4039.042	4150.936
C(pub_weekday)[T.Monday]	-54.6201	37.349	-1.462	0.144	-127.827	18.587
C(pub_weekday)[T.Saturday]	297.7369	111.740	2.665	0.008	78.716	516.758
C(pub_weekday)[T.Sunday]	275.6835	53.319	5.170	0.000	171.173	380.194
C(pub_weekday)[T.Thursday]	-58.3908	37.353	-1.563	0.118	-131.607	14.825
C(pub_weekday)[T.Tuesday]	-102.0492	36.247	-2.815	0.005	-173.096	-31.002
C(pub_weekday)[T.Wednesday]	-57.0209	37.214	-1.532	0.125	-129.964	15.922

Omnibus:	9722.016	Durbin-Watson:	1.809
Prob(Omnibus):	0.000	Jarque-Bera (JB):	120619.006
Skew:	2.266	Prob(JB):	0.00
Kurtosis:	14.696	Cond. No.	11.7

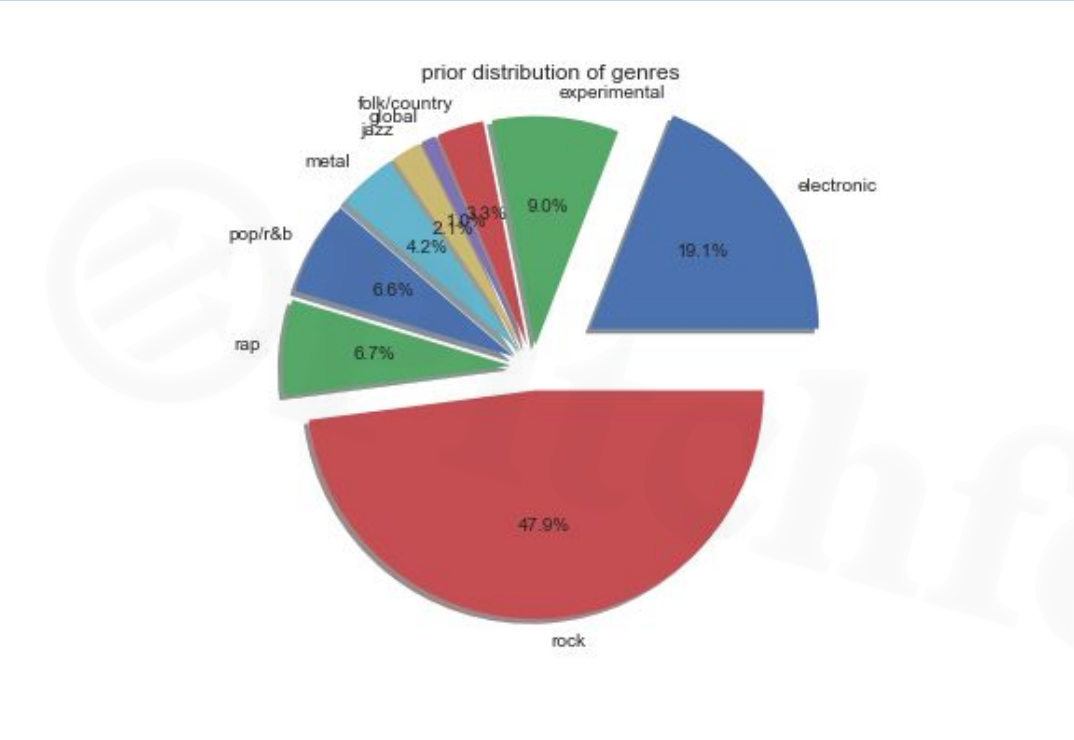
Conclusion:

Based on test results, we reject a null-hypothesis and conclude that with 5% level of confidence, review publication day effects content length.

	Degrees of Freedom	Sum of Squares	Mean Squares	F-Test	P-Value
C(pub_weekday)	6.0	1.429448e+08	23824140.0	11.6645	4.36447e-13
Residual	118394.0	3.756875e+10	2042446.0	NaN	NaN

Test #4: Chi-squared test (Goodness-of-fit)

Question: Is there any difference in distributions of ratings in past 2 years?



Goal:
Test, at 5 % level of significance, whether there is a sufficient evidence in the sample to conclude that the distribution of genres reviewed in magazine in the last 2 years had changed from previous years

Hypothesis:
- H0 = distribution of genres in the last two years ***has not*** changed
- HA = distribution of genres ***has*** changed

Chi-squared statistics = 356.813 and Chi-squared at alpha= 0.05 with 8 degrees of freedom = 15.507

Conclusion:

Based on test results, we reject a null-hypothesis and conclude that at 5% level of confidence there is a sufficient evidence of changing in distribution of genres reviewed in that magazine in the last two years