Sid Moturi
Ian Smith
Nive Venkat
Prutha Patel

1. Build a demand estimation regression model to estimate the drivers of app rank

We estimated the drivers of app rank using a Negative Binomial Regression (nbreg) model. The model was chosen because app rank is a discrete count variable and exhibited overdispersion (variance > mean), making nbreg more suitable than a Poisson model.

Model Specification

Dependent Variable: App Rank
Independent Variables: Price, File Size, Number of Screenshots, Rating Count, Average Rating, In-App Ad Dummy, In-App Purchase Dummy

Significance Threshold: Any variable with a p-value < 0.1 is considered significant

```
Negative binomial regression                    Number of obs     =     25,124
                                                LR chi2(7)        =     337.56
Dispersion     = mean                           Prob > chi2       =     0.0000
Log likelihood = -154393.48                     Pseudo R2         =     0.0011

------------------------------------------------------------------------------------
              rank |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------------+----------------------------------------------------------------
             price |   .0023886   .0014174     1.69   0.092    -.0003894    .0051666
          filesize |  -.0000966   .0000262    -3.69   0.000    -.0001479   -.0000453
    num_screenshot |   .0046961   .0020167     2.33   0.020     .0007434    .0086488
      rating_count |  -6.75e-07   3.96e-08   -17.07   0.000    -7.53e-07   -5.98e-07
    average_rating |   -.023766   .0056044    -4.24   0.000    -.0347503   -.0127817
     inapp_addummy |  -.0165306   .0108367    -1.53   0.127    -.0377701    .0047089
inapp_purchasedummy|  -.0804818    .010513    -7.66   0.000     -.101087   -.0598767
             _cons |   5.342026   .0262088   203.83   0.000     5.290658    5.393394
-------------------+----------------------------------------------------------------
          /lnalpha |  -.5240395   .0084171                     -.5405367   -.5075424
-------------------+----------------------------------------------------------------
             alpha |   .5921238    .004984                      .5824356    .6019732
------------------------------------------------------------------------------------
LR test of alpha=0: chibar2(01) = 1.9e+06              Prob >= chibar2 = 0.000
```

Regression Results and Interpretation:

Variable	Statistically Significant?	Coefficient	Interpretation
Price	Yes (p = 0.092)	0.0023886	The coefficient is positive, indicating that as the price increases, the Rank worsens (i.e., the rank increases numerically). This represents a negative effect, as a lower rank is more desirable.
filessize	Yes (p = 0.000)	-.0000966	The coefficient is negative, meaning that as the file size increases, the Rank improves (i.e., the rank decreases numerically). This improvement could potentially be attributed to larger apps offering more features or higher quality.

Sid Moturi
Ian Smith
Nive Venkat
Prutha Patel

num screenshots	Yes (p = 0.020)	0. 0046961	The coefficient is positive, indicating that a higher number of screenshots is associated with a worse (higher) rank. This is a negative effect, as the ideal rank is 1.
rating_count	Yes (p = 0.000)	-6.75e-07	The coefficient is negative, indicating that a higher rating count is associated with a lower (better) Rank, which is a positive effect on the ranking.
Average_rating	Yes (p = 0.000)	-.023766	The coefficient is negative, indicating that a higher average rating is associated with a lower (better) Rank.
inapp_addummy	No (p = 0.127, not significant)	-.0165306	The coefficient is negative and suggests that apps with in-app ads (when the variable is 1) tend to have a slightly better Rank (lower rank).
inapp_purchase dummy	Yes (p = 0.000)	-.0804818	This coefficient is negative and indicates that including in app purchase options leads to Slightly better rank (lower tank)

Key takeaways:

- The most important drivers of better rankings:
    - Higher rating count and average rating significantly improve rankings.
    - Larger file sizes are correlated with better ranks, suggesting users may perceive bigger apps as more feature-rich.
    - In-app purchases improve rankings, likely due to better engagement.
- Factors that negatively impact ranking:
    - Higher prices lead to worse rankings, though the effect is weak.
    - More screenshots are linked to worse rankings, which is unexpected.

Sid Moturi
Ian Smith
Nive Venkat
Prutha Patel

2. Compute the price elasticity of the app for Apple vs Google

We computed the price elasticity of the app for both Apple and Google using a standard regression (reg) model. The model was chosen because log-log regression models are well-suited for estimating elasticities. In the regression model, the coefficients of independent variables represent elasticities. For a 1% change in the independent variable, the elasticity tells us the percentage change in rank.

Model Specification

Dependent Variable: logrank
Independent Variables: logprice, logFilesize, logScreenshots, logRatingCount, average_rating, inapp_addummy, inapp_purchasedummy

Significance Threshold: Any variable with a p-value < 0.1 is considered significant

Google

Source	SS	df	MS
Model	1125.64263	7	160.80609
Residual	20086.767	21,531	.932923086
Total	21212.4096	21,538	.984882979

Number of obs	=	21,539
F(7, 21531)	=	172.37
Prob > F	=	0.0000
R-squared	=	0.0531
Adj R-squared	=	0.0528
Root MSE	=	.96588

logrank	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
logprice	.0175327	.0105564	1.66	0.097	-.0031586	.0382239
logFilesize	-.0088381	.0059093	-1.50	0.135	-.0204207	.0027446
logScreenshots	.1219981	.0186631	6.54	0.000	.085417	.1585792
logRatingCount	-.1017773	.0033645	-30.25	0.000	-.108372	-.0951826
average_rating	.0795643	.0080702	9.86	0.000	.0637462	.0953824
inapp_addummy	.011486	.0151419	0.76	0.448	-.0181933	.0411653
inapp_purchasedummy	-.1462966	.0153173	-9.55	0.000	-.1763197	-.1162735
_cons	5.050585	.0427848	118.05	0.000	4.966724	5.134447

Apple

Source	SS	df	MS
Model	281.276101	7	40.1823001
Residual	3314.54325	3,535	.937635997
Total	3595.81935	3,542	1.01519462

Number of obs	=	3,543
F(7, 3535)	=	42.85
Prob > F	=	0.0000
R-squared	=	0.0782
Adj R-squared	=	0.0764
Root MSE	=	.96832

logrank	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
logprice	-.0621665	.0324234	-1.92	0.055	-.1257368	.0014039
logFilesize	-.0883939	.0157791	-5.60	0.000	-.119331	-.0574567
logScreenshots	-.1284875	.0543716	-2.36	0.018	-.2350903	-.0218847
logRatingCount	-.1226005	.0085118	-14.40	0.000	-.1392891	-.1059118
average_rating	.1599435	.0516807	3.09	0.002	.0586165	.2612705
inapp_addummy	.0896564	.035297	2.54	0.011	.0204518	.158861
inapp_purchasedummy	.0384516	.0387482	0.99	0.321	-.0375194	.1144226
_cons	5.720704	.2236447	25.58	0.000	5.282219	6.15919

Regression Results and Interpretation:

Variable	Google		Apple		Interpretation
	Coefficient	*Significance*	*Coefficient*	*Significance*	
log(price)	0.0175327	Yes (p = 0.097)	-0.0621665	Yes (p = 0.055)	Higher price deters Google users leading to a worse rank, while Apple users are likely to download the app
log(file size)	-0.0088381	No (p = 0.135)	-0.0883939	Yes (p = 0.000)	Larger file size leads to increased downloads for Apple users, while there is no statistically significant evidence that it affects Google users
log(screenshots)	0.1219981	Yes (p = 0.000)	-0.1284875	Yes (p = 0.018)	More screenshots increases the likelihood that Apple users would download the app, while Google users would be deterred

Sid Moturi
Ian Smith
Nive Venkat
Prutha Patel

log(rating count)	-0.1017773	Yes (p = 0.000)	-0.1226005	Yes (p = 0.000)	For both Apple and Google, more ratings increase the likelihood of downloads
Average Rating	0.0795643	Yes (p = 0.000)	0.1599435	Yes (p = 0.002)	Increasing the average rating would deter both Apple and Google users
In-App Ad Dummy	0.011486	No (p = 0.448)	0.0896564	Yes (p = 0.011)	Showing in-app ads deters Apple users from downloading the app, while there is no statistically significant evidence that it affects Google users
In-App Purchase Dummy	-0.1462966	Yes (p = 0.000)	0.0384516	No (p = 0.321)	In-app purchases likely encourages Google users, while there is no statistically significant evidence that it affects Apple users

Key Takeaways

- **Google Model**: Visual presentation using screenshots and user feedback (i.e. rating count and average rating) were the most significant drivers
- **Apple Model**: Variables related to app size, ratings, and overall quality (i.e. file size, ratings count, average rating) had stronger significance
- The rating count was strongly significant for both the Apple and Google models

NYU STERN

Sid Moturi
Ian Smith
Nive Venkat
Prutha Patel

3. **Compute the price elasticity of the app for US Vs China and Tablet Vs Smartphone**
    a. **Comparison of Price Elasticity for the US vs. China**

We estimated the price elasticity of demand for mobile apps separately for the US (rindex == 1) and China (rindex == 2).

China

Source	SS	df	MS
Model	630.489976	7	90.0699966
Residual	10631.0301	10,580	1.00482326
Total	11261.5201	10,587	1.06371211

Number of obs	=	10,588
F(7, 10580)	=	89.64
Prob > F	=	0.0000
R-squared	=	0.0560
Adj R-squared	=	0.0554
Root MSE	=	1.0024

logrank	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
logprice	.016466	.0158048	1.04	0.298	-.0145143	.0474463
logFilesize	-.0226322	.0086069	-2.63	0.009	-.0395033	-.0057611
logScreenshots	.0413288	.0274865	1.50	0.133	-.0125499	.0952075
logRatingCount	-.1133302	.005028	-22.54	0.000	-.123186	-.1034745
average_rating	.093437	.01828	5.11	0.000	.0576049	.1292692
inapp_addummy	.0729783	.0219106	3.33	0.001	.0300293	.1159272
inapp_purchasedummy	-.125985	.0220379	-5.72	0.000	-.1691834	-.0827866
_cons	5.333665	.0908044	58.74	0.000	5.155671	5.511658

USA

Source	SS	df	MS
Model	466.212414	7	66.6017734
Residual	13043.9978	14,528	.897852271
Total	13510.2102	14,535	.929495027

Number of obs	=	14,536
F(7, 14528)	=	74.18
Prob > F	=	0.0000
R-squared	=	0.0345
Adj R-squared	=	0.0340
Root MSE	=	.94755

logrank	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
logprice	.0095753	.0139832	0.68	0.494	-.0178336	.0369841
logFilesize	-.0353542	.0070548	-5.01	0.000	-.0491824	-.021526
logScreenshots	.0277856	.0225026	1.23	0.217	-.0163223	.0718935
logRatingCount	-.0701853	.0038372	-18.29	0.000	-.0777067	-.0626639
average_rating	.0540669	.0089517	6.04	0.000	.0365205	.0716133
inapp_addummy	.0237914	.0181191	1.31	0.189	-.0117243	.0593071
inapp_purchasedummy	-.1319997	.0190457	-6.93	0.000	-.1693318	-.0946676
_cons	5.247716	.0470244	111.60	0.000	5.155542	5.33989

<u>Model Specification</u>

Dependent Variable: logrank
Independent Variables: logprice, logFilesize, logScreenshots, logRatingCount, average_rating, inapp_addummy, inapp_purchasedummy

Significance Threshold: Any variable with a p-value < 0.1 is considered significant

Variable	USA		China		Findings
	Coefficient	*Significance*	*Coefficient*	*Significance*	
log(price)	0.0096	No (p = 0.494)	0.0165	No (p = 0.298)	Price does not significantly impact rank in either market.
log(file size)	-0.0354	Yes (p = 0.000)	-0.0226	Yes (p = 0.009)	Larger file sizes improve rank in both countries, but more so in the US.
log(screenshots)	0.0278	No (p = 0.217)	0.0413	No (p = 0.133)	Number of screenshots does not significantly impact rank.
log(rating count)	-0.0701	Yes (p = 0.000)	-0.1133	Yes (p = 0.000)	Higher rating counts improve rank significantly in both markets, with a stronger effect in China.
Average Rating	-0.0541	Yes (p = 0.000)	-0.0934	Yes (p = 0.000)	Higher average ratings improve rank in both markets, with a stronger effect in China.
In-App Ad Dummy	0.0238	No (p = 0.189)	0.0729	Yes (p = 0.001)	In-app ads improve rank in China but have no significant effect in the US.

Sid Moturi
Ian Smith
Nive Venkat
Prutha Patel

In-App purchase	-0.1319	Yes (p = 0.000)	-0.1260	Yes (p = 0.000)	In-app purchases improve rank significantly in both markets, slightly more in the US.

Key takeaways: USA vs China

- Ratings and rating count are the strongest drivers of rank in both the US and China, but their impact is stronger in China.
- File size positively affects rank in both markets, but the effect is stronger in the US.
- In-app purchases significantly improve rank in both markets, slightly more in the US.
- In-app ads significantly improve rank in China, but not in the US.
- Price does not significantly impact rank in either market.
- Number of screenshots does not significantly impact rank in either market.

Comparison of Price Elasticity for the Tablet vs. Smartphone

We estimated the price elasticity of demand for mobile apps separately for the smartphone (deviceindex == 1) and Tablet (deviceindex == 2).

Smartphone

Source	SS	df	MS
Model	625.767004	7	89.3952863
Residual	13987.629	14,671	.953420282
Total	14613.396	14,678	.995598581

Number of obs =	14,679
F(7, 14671) =	93.76
Prob > F =	0.0000
R-squared =	0.0428
Adj R-squared =	0.0424
Root MSE =	.97643

logrank	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
logprice	.0390763	.0142589	2.74	0.006	.0111272	.0670255
logfilesize	-.0185958	.0073926	-2.52	0.012	-.0330862	-.0041054
logscreenshots	-.0135873	.0226547	-0.60	0.549	-.0579933	.0308187
logratingcount	-.0702542	.0032863	-21.38	0.000	-.0766957	-.0638126
average_rating	.0724599	.0105313	6.88	0.000	.0518172	.0931026
inapp_addummy	.0676122	.0180544	3.74	0.000	.0322233	.1030012
inapp_purchasedummy	-.1951497	.0188581	-10.35	0.000	-.232114	-.1581854
_cons	5.204497	.055041	94.56	0.000	5.09661	5.312384

Tablet

Source	SS	df	MS
Model	690.832578	7	98.6903683
Residual	9600.34099	10,437	.919837212
Total	10291.1736	10,444	.985367059

Number of obs =	10,445
F(7, 10437) =	107.29
Prob > F =	0.0000
R-squared =	0.0671
Adj R-squared =	0.0665
Root MSE =	.95908

logrank	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
logprice	-.0017874	.0141056	-0.13	0.899	-.029437	.0258622
logfilesize	-.0309022	.0081945	-3.77	0.000	-.0469649	-.0148395
logscreenshots	.1801224	.0271537	6.63	0.000	.1268958	.2333489
logratingcount	-.1239387	.004953	-25.02	0.000	-.1336476	-.1142299
average_rating	.0834297	.0118676	7.03	0.000	.060167	.1066924
inapp_addummy	-.002623	.0217137	-0.12	0.904	-.0451861	.0399401
inapp_purchasedummy	-.0289564	.0218169	-1.33	0.184	-.0717217	.013809
_cons	5.084598	.0620603	80.63	0.000	4.960987	5.208208

Sid Moturi
Ian Smith
Nive Venkat
Prutha Patel

Model Specification

Dependent Variable: logrank
Independent Variables: logprice, logFilesize, logScreenshots, logRatingCount, average_rating, inapp_addummy, inapp_purchasedummy

Significance Threshold: Any variable with a p-value < 0.1 is considered significant

Variable	Smartphone		Tablet		Findings
	Coefficient	*Significance*	*Coefficient*	*Significance*	
log(price)	.039	yes (p = 0.006)	-.001787	No (p = 0.899)	Price does not significantly impact rank in Tablets but does have an impact with smartphones
log(file size)	-.018595	Yes (p = 0.012)	-.0309022	Yes (p = 0.000)	Larger file sizes improve rank for both Tablet and Smartphone
log(screenshots)	-.013587	No (p = 0.549)	.1801224	Yes (p = 0.000)	Number of screenshots does not significantly impact Smartphone but does impact Tablet
log(rating count)	-.0702542	Yes (p = 0.000)	-.1239387	Yes (p = 0.000)	Higher rating counts improve rank significantly with both Smartphone and Tablet
Average Rating	-.0724599	Yes (p = 0.000)	.0834297	Yes (p = 0.000)	Higher average ratings improve rank with both Smartphone and Tablet
In-App Ad Dummy	.0676122	yes (p = 0.000)	-.002623	No (p = 0.984)	In-app ads impact Smartphones but do not impact tablets.
In-App purchase	-.1951497	Yes (p = 0.000)	-.0289564	No (p = 0.184)	In-app purchases improve rank significantly with Smartphone but is not a significant driver for Tablets

Key takeaways: Smartphones Vs. Tablets

- Ratings and rating count are the strongest drivers of rank for both Smartphones and Tablets
- File size has a greater impact on rank for tablets than it does with smartphones

Sid Moturi
Ian Smith
Nive Venkat
Prutha Patel

- In-app purchases significantly improves rank in Smartphones than it does with tablets
- In-app ads significantly impacts rank on Smartphones
- Price does not significantly impact rank in Tablets but does have an impact with smartphones
- Number of screenshots does not impact rank within smartphones but is a significant driver for rank within tablets

----