Oh, Andrea. The Battle of the Book Cover/ https://tinyurl.com/y5hdwv9a

Goal: Suggestions or improvements

We propose using modern machine learning techniques to learn successful covers for books.

By doing so, we aim to discover whether certain elements of book covers are related to the eventual success of a book.

More importantly, knowing this information, can we generate suggestions to improve cover design?

Can we actually judge a book by its cover?

Schedule

Week 1: gather data, exploration, visualization, scrape covers

Week 2: given a book cover, get related book covers

Week 3: simple web service to gather related book covers

Week 4: CNN model for feature extraction, augment related book cover

Week 5: Feature extraction, suggestion service

Week 6: Profit

Review/Metadata Extraction

Given book reviews and Amazon metadata:

- Consolidate reviews to get an overall rating weighted by number of reviews
- Extract book rankings and also bought books for relationships

Ratings

Job information | Results | JSON | Execution details

Row	total_reviews	rating	ASIN
1	115455.0	3.984641932700604	038568231X
2	91711.0	4.105053489100756	B000X1MX7E
3	88491.0	4.76244550885313	0312577222
4	87513.0	4.683847141939627	B003156C4E
5	87279.0	4.683857464849201	B001C4VLZQ

Rankings

Row	ASIN	ranking	also_bought
1	B01HI5V3HI	3921775	
2	B01HIP8KLU	2655897	
3	B01HJ0AYR2	583884	B078V7G3XY,B01M0C8BVU,B0747HDXCX,B00L43QLD2,B00VEE42DW,B012JPGK3E,B01MR8C5PW,B01EH10DAE,B009B2YKTU,B00L1BBW4A,B00RHCX25O,B005T54LCY,B00MMNTO14,B014OL8SG8,B00SQ4DLBM,B00IFSU3OM,B009FR9CCG,B01L0Y8P2O,B01B5BMVXQ,B00CENJB5G,B01L0Y8P1A,B008MM8SBG,B00DPZZVBU,B07DMZB97R,B00FD9WT2A,B07CQ7P24Y,B009CQG9T4,B00KP8DA72,B07JWFDM2F,B004RKXHEU,B07HDMC4GH,B07KGDDW7J,B07B26JHMD,B07FCQV31V,B077WXP3KG,B076X2TNRM,B002G54Y2M,B06XCM8VPB,B079KX1XFD,B078CY7TLY,B078X12Y75,B074N1XJR2,B07BR6RNS3,B07BK3M7V7,B01EROMI1S,B074PD2QR3,B07D4M7N3L,B071X2K4S6,B075C9TMR1,B002361LHU,B002DW92V6,B01H03IBHS,B01MRF794N,B00OBT9B60,B0783P29QP,B01NCU3JYC,B00UI4ERXY,B075JGPRR5,B07634GR5R,B000OCXHT0,B01L9W8CYI,B00KWFZ7CE
4	B00T3PMFHO	1103611	B004G8PIOA,B007WKFMGS
5	B00T884AME	126974	B01D6RFQPC,B01NC30DNW,B07652YGR9,B01CNZV0Q2,B01BIFXV88,B07CK1MRPJ,B00PMVTNQO,B01N3QYKNX,B073W9DBPS,B075KS2V9N,B07BBX3FMN,B01A86GUTK,B07CZBV5LQ,B07K4TS4SR,B00KECZLG2,B01N6SFVMS,B00I5Q71YE,B07D67FG2H,B010XQ4WDG,B01FIJKZNY,B00LKGBK2A,B01MY9ZP2W,B07K1FJ75F,B077X4LQ1S,B01MYC4RV7,B07FZTSFXY,B00ZJG398U,B01NAB6G4W,B013CFG1HA,B072L1NZ16,B07GHYQ4T1,B008Y0XH1C,B00WHXYZG8,B01BECUPJS,B0793X6RJQ,B01A7RJ94S,B00495Z8E0,B07HLC6R6L,B077T898H9,B01LVXR8M8,B07L5R5KWZ,B076HKP2PM,B07D8JM9W9,B0153IIWDK,B00NCYZ2M2,B0166SHTVM,B00OO5IN1A,B07BX14NRW,B01JX6EFKW,B00EBC0ZFI,B074WXMKVJ,B008ZEDO06,B017RMH4GG,B07CZNC58K,B078RQNCDT,B00MX0EC8G,B01MT3UWAM,B005SGCD6A,B00CFQJNKA,B00LB6DKD6,B000O2U9AFA,B008H7CEQG,B019SK5IRW,B000OVLIQU,B01DLBGDSC,B005FVPFEK,B01N27XYYO,B07FH84SVH,B075LYJWN5,B01FUV0WJW,B00JUUZOD4,B00HGB70WM,B00YQKID04,B01M662J9X,B0130L

Methods

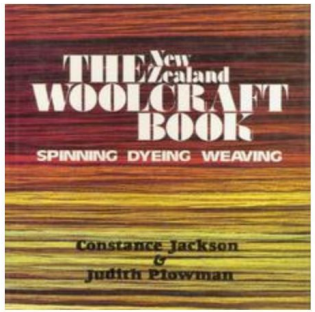
1. Data collection and cleaning

2. Image feature extraction
   a. Traditional image processing to extract predefined features
   b. CNN
3. Recommendation generation
   a. K-nn
   b. Decision tree regression

Image Feature Extraction

Three simple image features extracted for MVP

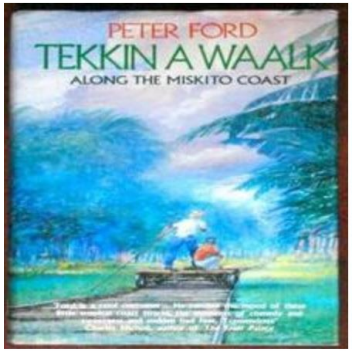
1. Dominant Color
   a. K-means clustering on pixel RGB values to group top 5 colors. Choose mean of largest cluster
2. Brightness
   a. RMS of all the pixel values
3. Colorfulness
   a. Metric evaluating the variation in RGB values



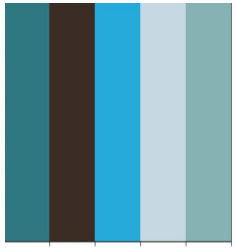
Top 5 Colors

Colorfullness
65.36

Brightness
119.04



Top 5 Colors



Colorfullness
90.76

Brightness
158.97

System

- All code managed in GitHub
- All coordination through Slack
- Feature data and metadata trained using Google Dataproc
- Really simple frontend (pure html)
- Flask app for backend
- Docker images built on Google Cloud Build
- Deployed via Kubernetes on Google Kubernetes Engine
- Grabs images from Google Cloud Storage
- Will grab metadata from Google BigQuery

Demo!

http://35.222.73.123/

Thank you!

Charlie Summers
(charles.summers@columbia.edu)

Daniel Silva
(dcs2180@columbia.edu)

Najim Yaqubie
(najim.yaqubie@columbia.edu)