

ANNEX

Figure 1: Most Common News Categories

Most Common News Categories

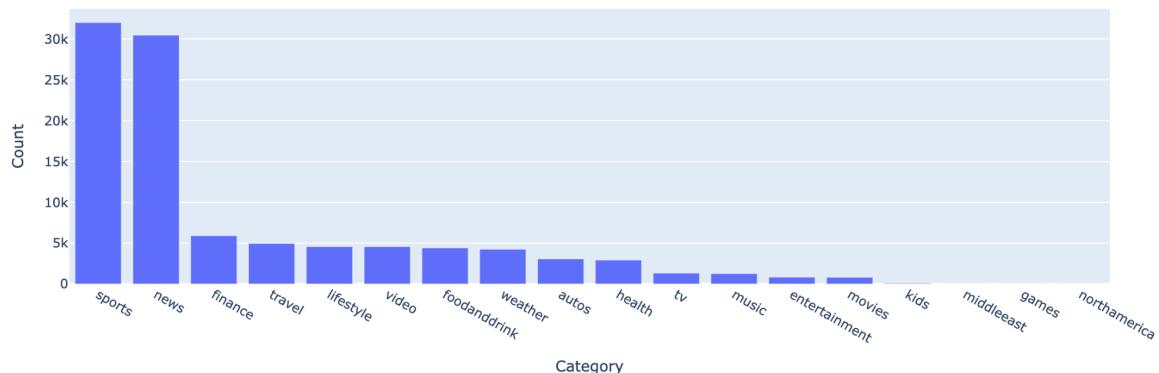


Figure 2: Treemap of Categories and Subcategories

Treemap of Categories and Subcategories

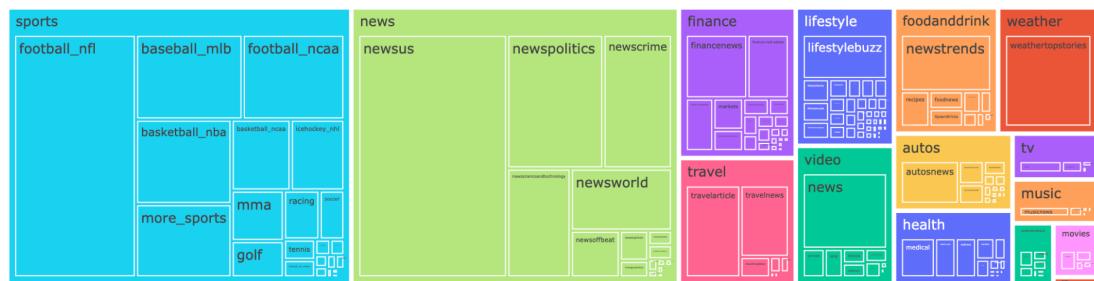


Figure 3: Subcategories for Sports

Subcategories for sports

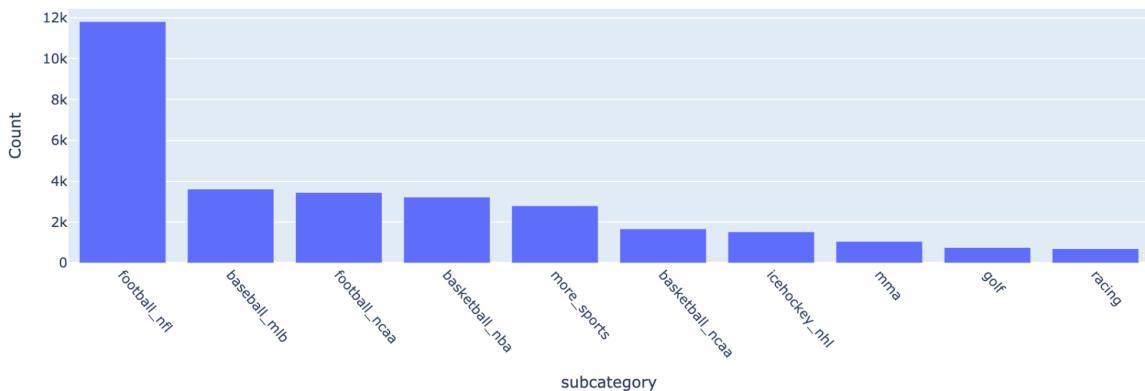


Figure 4: Subcategories for News

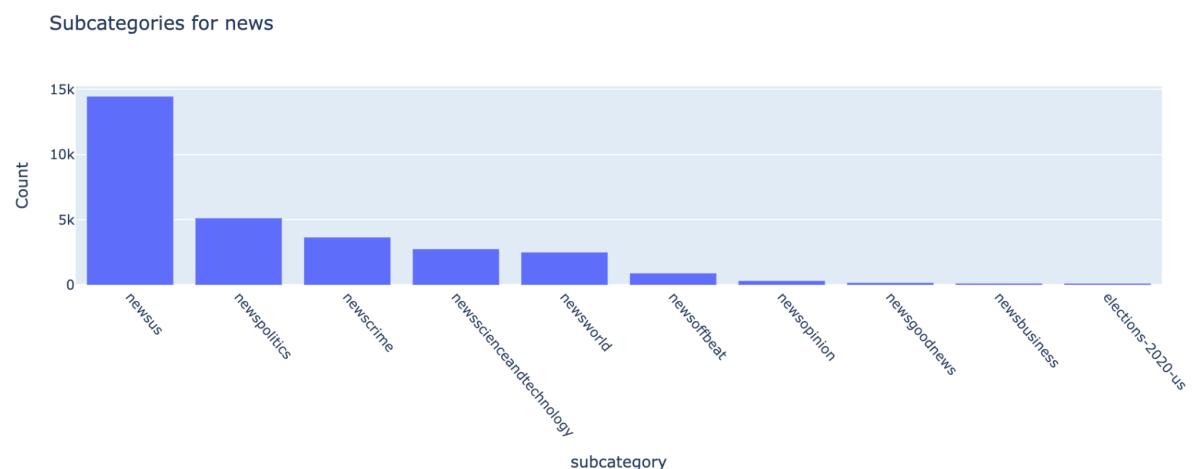


Figure 5: Average Title Word Count per Category

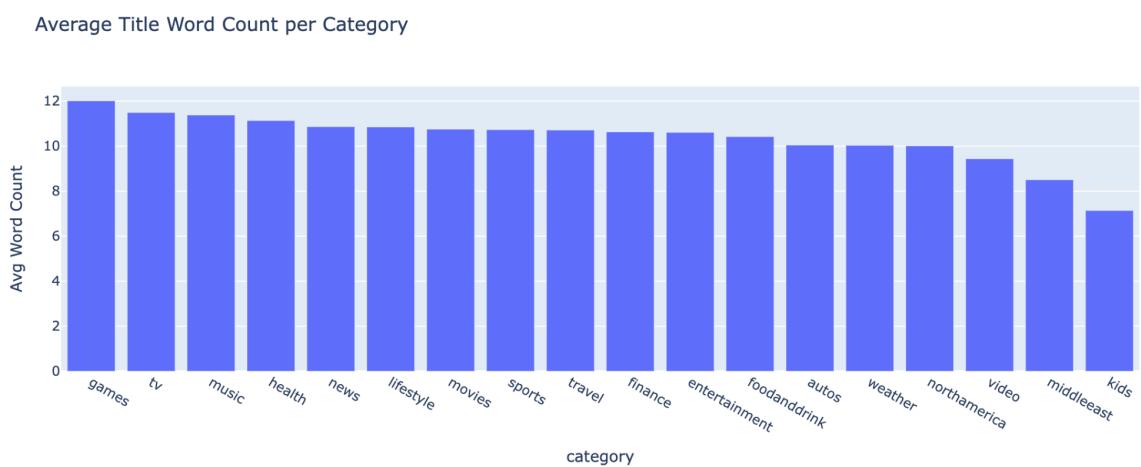


Figure 6: Average Title Polarity per Category

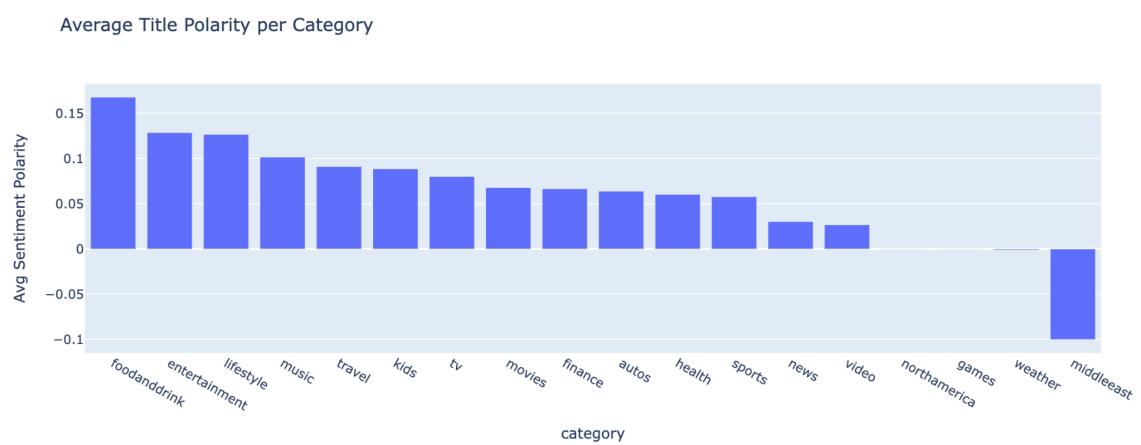


Figure 7: Distribution of Title Features

Distribution of Title Features

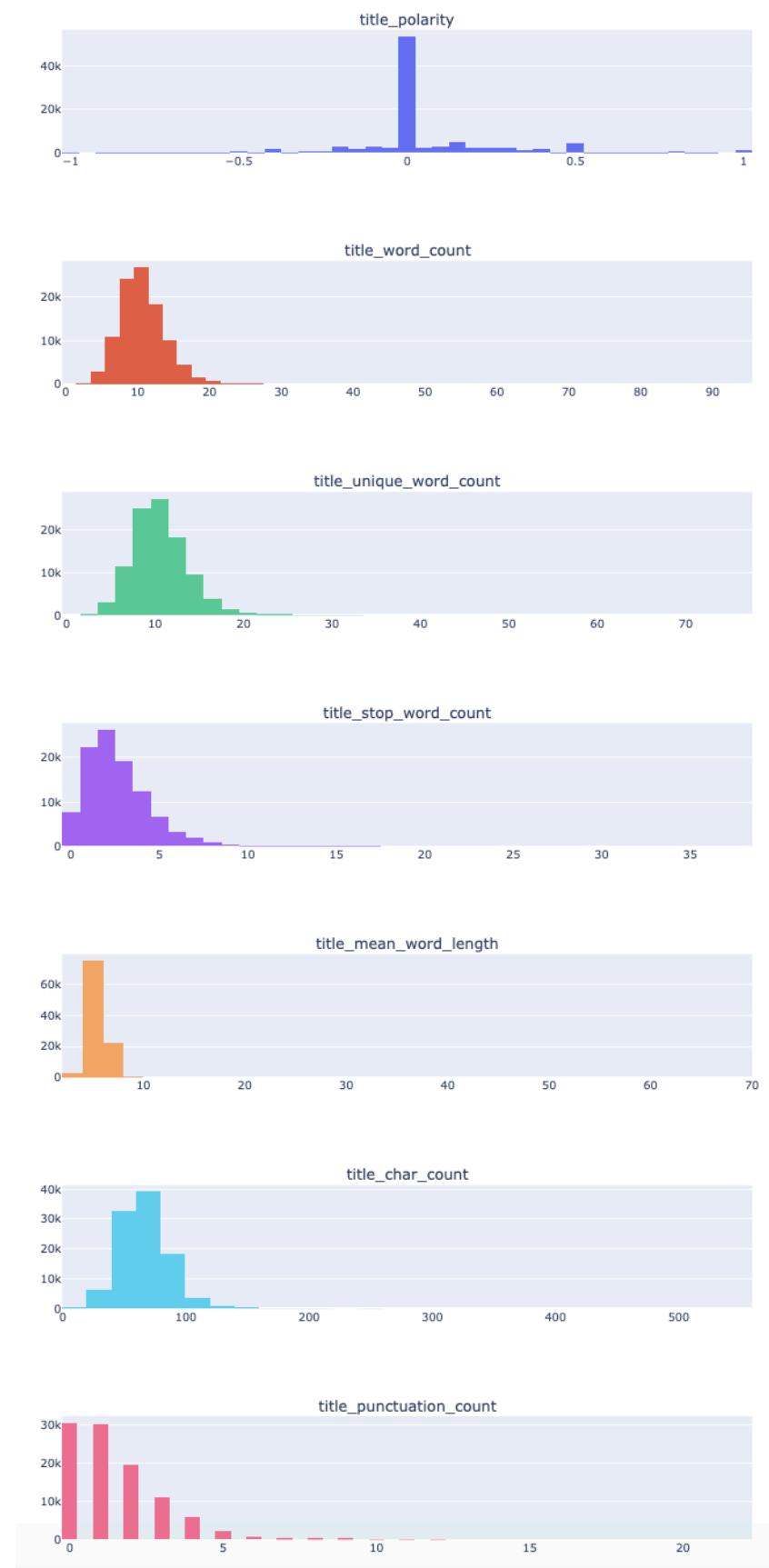


Figure 8: Most Frequent Words in News Titles

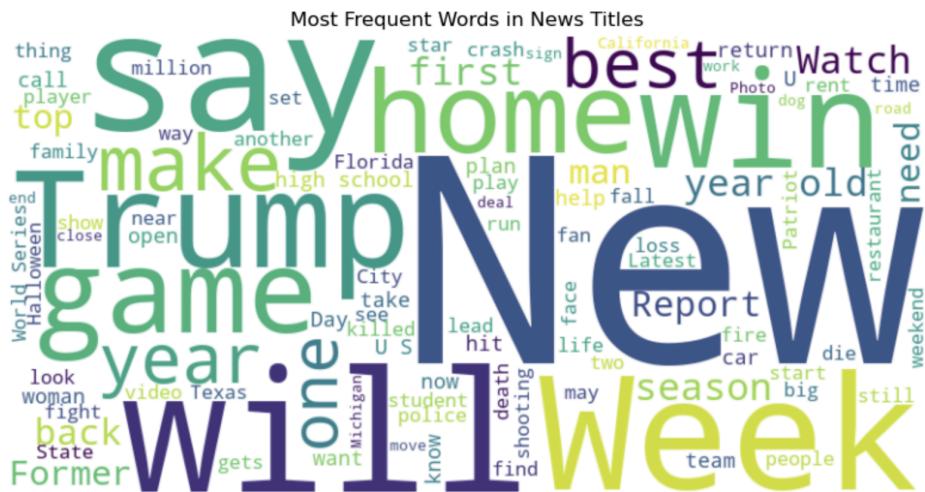


Figure 9: Most Frequent Words in Categories



Figure 10: Distribution of News Abstract Length

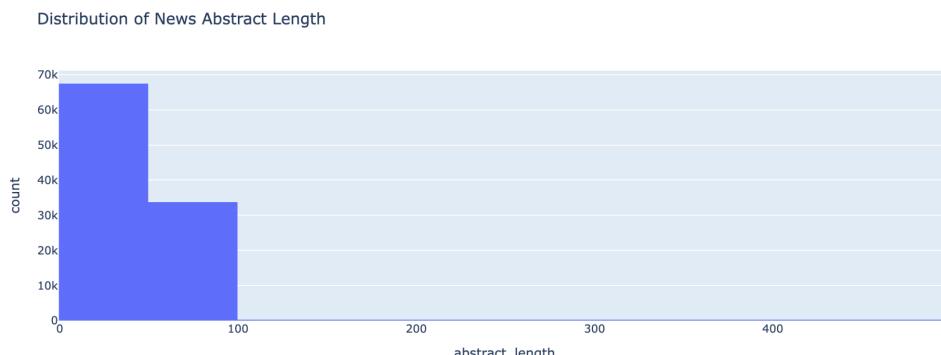


Figure 11: Average Abstract Length by Category

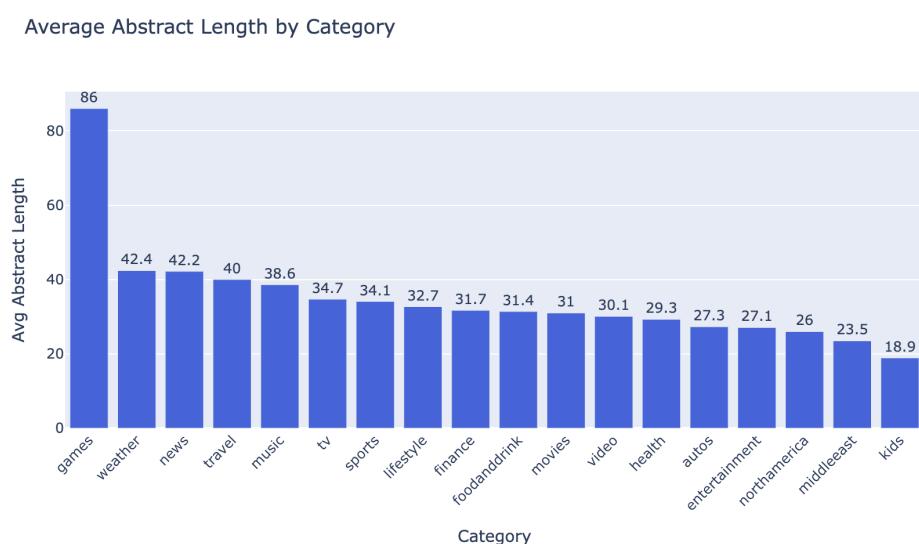


Figure 12: Distribution of Abstract Features



Figure 13: Average Abstract Word Count per Category

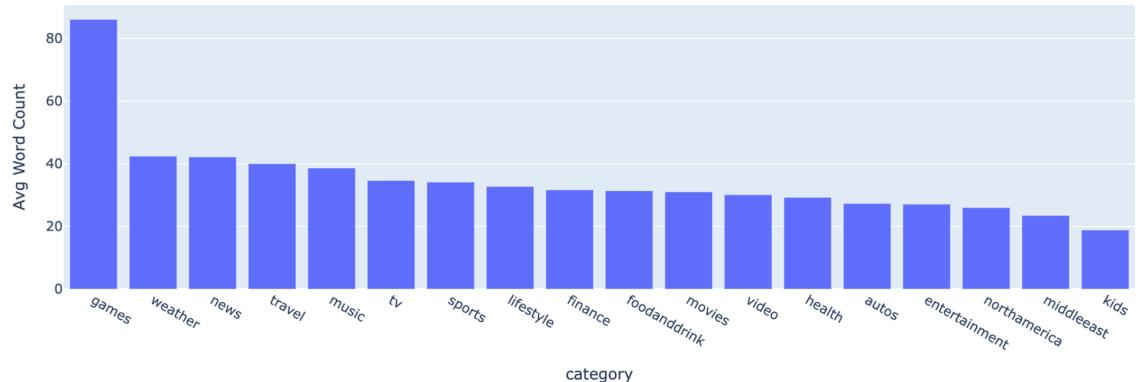


Figure 14: Average Abstract Polarity per Category



Average Abstract Polarity per Category

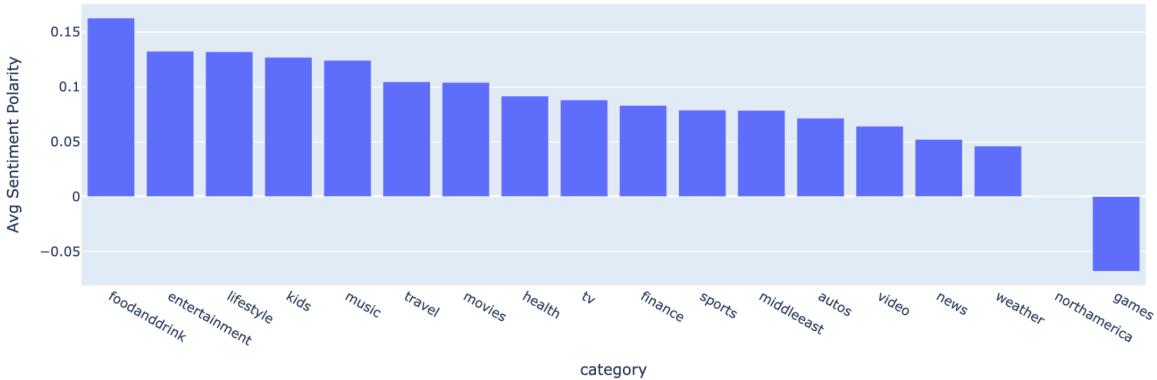


Figure 15: Most Frequent Words in News Abstracts



Figure 16: Most Frequent Words in News Abstracts per Category



Figure 17: User Activity per Hour

User Activity by Hour

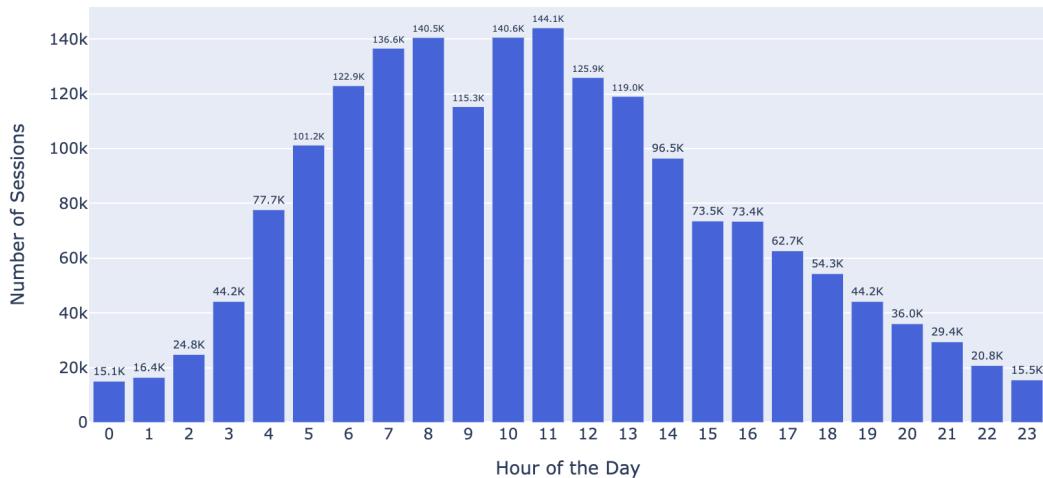


Figure 18: User Activity by Day of Week

User Activity by Day of the Week

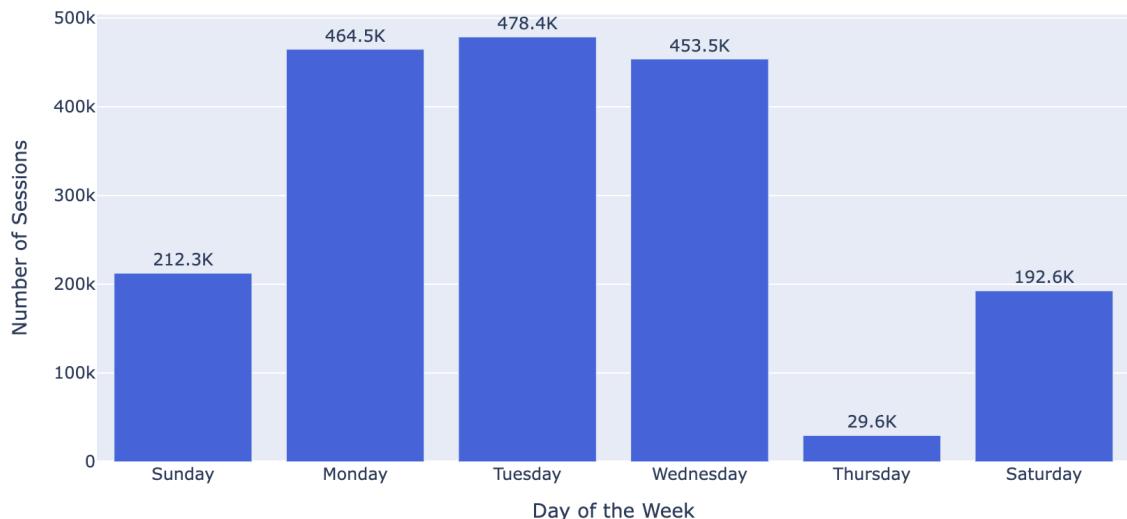


Figure 19: Distribution of User Activity

Distribution of User Activity (Number of Sessions per User)

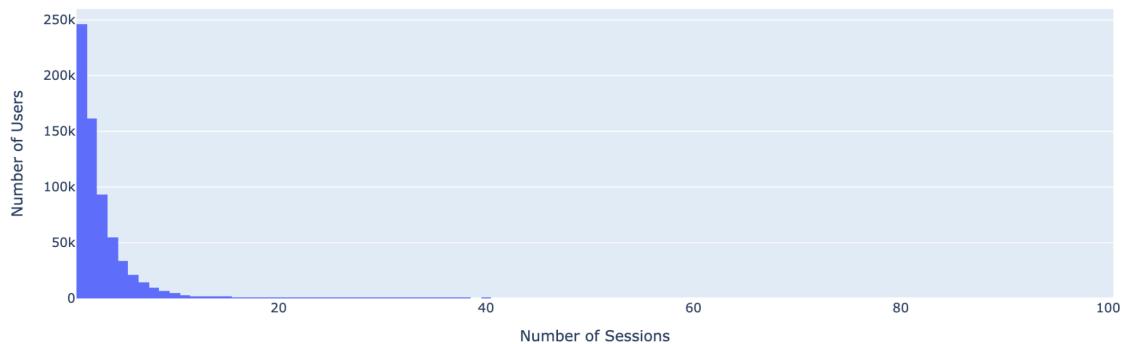


Figure 20: Distribution of History Length per User Session

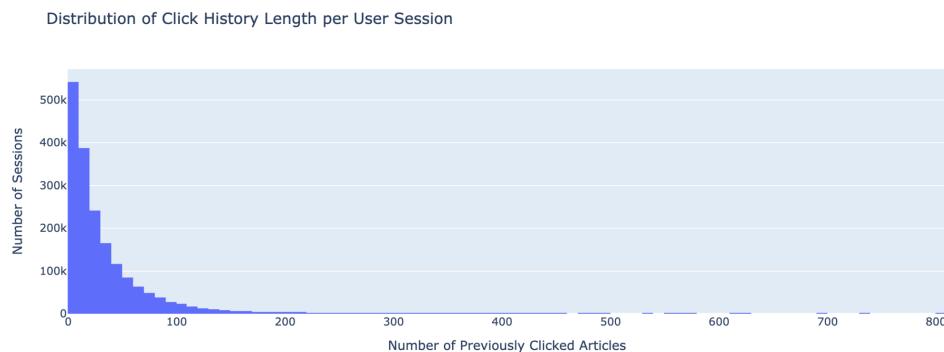


Figure 21: Distribution of Number of Clicks per User Session

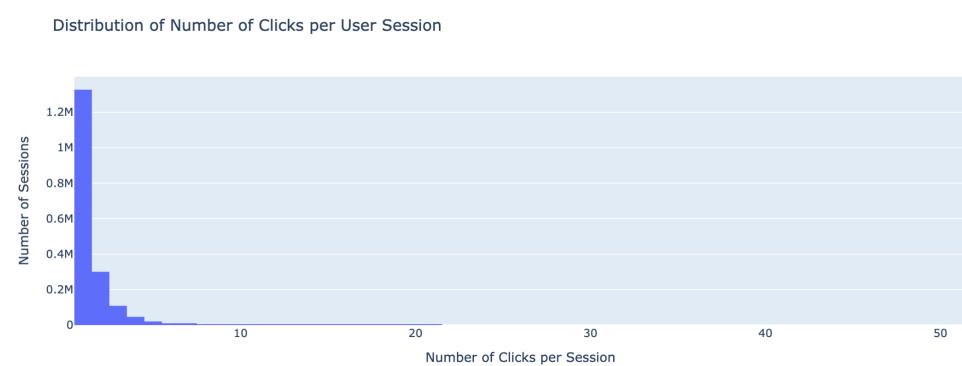


Figure 22: Top Entities Mentioned in News Articles

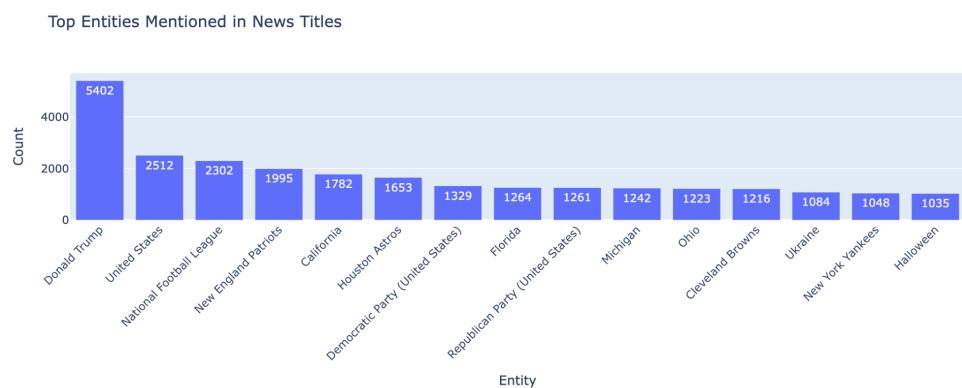


Figure 23: Average Number of Entities per News Article by Category

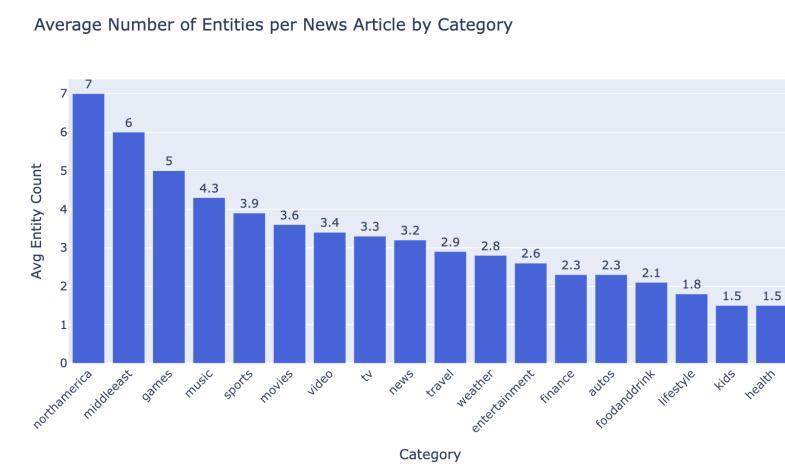
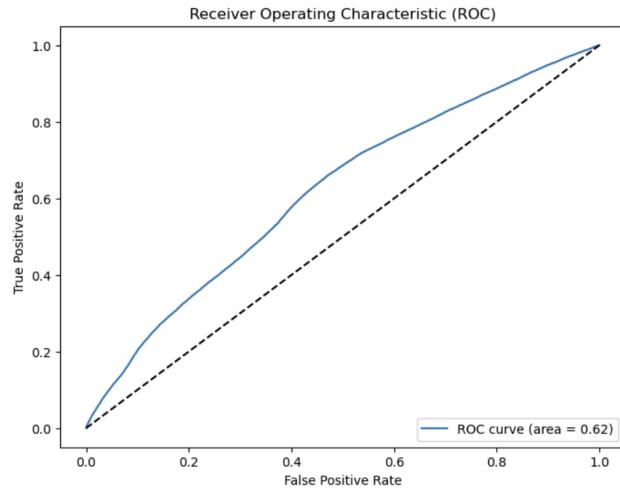


Figure 24: Light GBM Model Grid Search Results

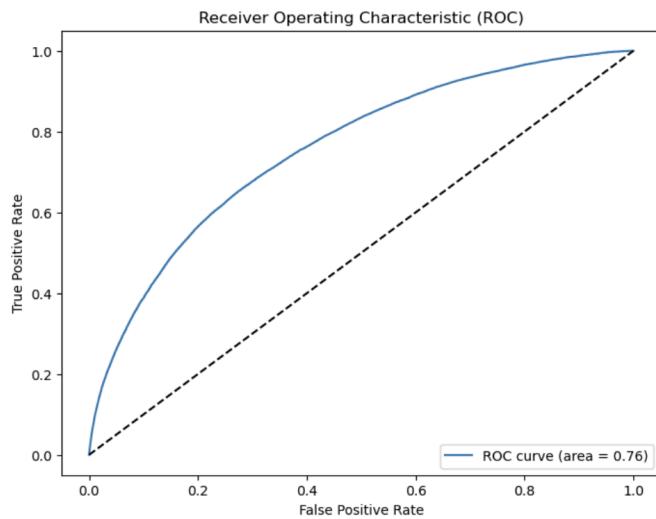
	mean_test_score	param_num_leaves	param_max_depth	param_learning_rate	param_n_estimators	param_min_child_samples
19	0.7556	63	-1	0.05	300	20
23	0.7553	63	-1	0.05	300	50
31	0.7547	63	10	0.05	300	50
27	0.7543	63	10	0.05	300	20
18	0.7507	31	-1	0.05	300	20
22	0.7506	31	-1	0.05	300	50
26	0.7501	31	10	0.05	300	20
30	0.7501	31	10	0.05	300	50
17	0.7469	63	-1	0.05	100	20
21	0.7469	63	-1	0.05	100	50

Figure 25: Logistic Regression Metric Performance



	AUC	Precision	Recall	MAP	nDCG	Precision@5	Recall@5	nDCG@5	Precision@10	Recall@10	nDCG@10
0	0.6185	0.0567	0.558	0.0636	0.539	0.1442	0.6428	0.4689	0.0957	0.8316	0.5334

Figure 26: Light GBM Metric Performance



	AUC	Precision	Recall	MAP	nDCG	Precision@5	Recall@5	nDCG@5	Precision@10	Recall@10	nDCG@10
0	0.7574	0.0878	0.667	0.1411	0.6299	0.1745	0.7646	0.587	0.1053	0.9025	0.6348

Table 1 - Impressions

Variables	Values
Total impressions in dataset (Nov 9-22, 2019)	4,979,946
% of year recorded in dataset	3.56%
Total impressions / year	139,821,561
% increase in impressions from recommender	50%
New impressions	69,910,780
Total impressions	209,732,341

Table 2 - CPM Model

Variables	# Values
Impressions	209,732,341.00
CTR Delta	0.14
Ads Per Page	5.00
eCPM (per 1000)	5.00
Ad Revenue	755,036.43

Table 3 - CPC Model

Variables	# Values
Impressions	209,732,341.00
CTR Delta	0.14
Ads Per Page	5.00
Ad CTR Delta	0.01
CPC (€)	0.60
Ad Revenue	906,043.71

Table 4 - Subscribers & New Subscribers

Variables	Values
Total users in dataset (Nov 9-22, 2019)	876,956
Days / dataset	13
Days / year	365
Percentage of users that return per 13-day cycle	40%
Cycles / year	28
Returning users	350,782
New users per cycle	526,174
Total users / year	15,124,118

Variables	Values
Total Users	15,124,118.00
New Conversion Rate	0.04
Old Conversion Rate	0.02
New Subscribers	302,482.00

Table 5 - Churn

Variables	Values
Churn rate before	10%
Churn rate after	5%
Churn reduction (subscriptions)	15,124
Churn reduction (€)	1,814,894

Table 6 - Subscription Revenue

Variables	# Values
New Subscribers	302,482
Annual Revenue per Subscriber	120
Subscription Revenue (€)	38,112,734

Table 7: Cloud Infrastructure

Name	Cost (€)	Total Cost (€)/ year
Serverless	4	559
Dedicated VMs	10	1,398
Standard NVidia A100 GPU VM (per hour)	3.06	26,806
Storage (1M)	0.28	470
Networking costs (Data Transfers and API Ca	0.08	1150
Total		30,383

Table 8: Model Training

Name	Cost	Total Cost	
ND96amsr A100 v4 (on demand/ hour)	32.77	25560.6	15 hours /week
Prototype (6M)		60000	
MVP (1Y)		120000	
Full Scale Deployment		400000	
Continuous Model Retraining		1,200,000	
Total		1805560.6	

Table 9: Salaries

Name	Average Salary	Team Size	Total Cost
Data Scientist	100000	5	500000
ML Engineer	110000	4	440000
Data Engineer	90000	3	270000
Cloud/ DevOps Engineer	120000	2	240000
Total			1450000

Table 10: Compliance & Security

Name	Cost
GDPR Initial Compliance	30000
GDPR Annual Maintenance	20000
AI Act Compliance per Model	52000
Cybersecurity Implementation	75000
Total	177000

Table 11: Cloud Infrastructure after 50% growth (impressions)

Name	Cost (€)	Total Cost (€)/ year
Serverless	4	839
Dedicated VMs	10	2,097
Standard NVidia A100 GPU VM (per hour)	3.06	26,806
Storage (1M)	0.28	705
Networking costs (Data Transfers and API Ca	0.08	1,594.08
Total		32,041

Table 12: Model Training after 50% growth

Name	Cost	Total Cost
ND96amsr A100 v4 (on demand/ hour)	32.77	38340.9
Prototype (6M)		60000
MVP (1Y)		120000
Full Scale Deployment		400000
Continuous Model Retraining		1,680,000
Total		2298340.9

Table 13: Salaries after 50% growth

Name	Average Salary	Team Size	Total Cost
Data Scientist	100000	6	600000
ML Engineer	110000	5	550000
Data Engineer	90000	4	360000
Cloud/ DevOps Engineer	120000	2	240000
Total			1750000

Table 14: Compliance & Security after 50% growth

Name	Cost
GDPR Initial Compliance	30000
GDPR Annual Maintenance	30000
AI Act Compliance per Model	65000
Cybersecurity Implementation	120000
Total	245000

Table 15: ROI

Year	Incr. Ad Revenue	Incr. Subscription Revenue	Total Incr. Revenue	Investment	#	ROI
1	€1,661,080.14	€38,112,734.16	€39,773,814.30	€4,325,381.53	8.20	
2	€1,827,188.15	€41,924,007.58	€43,751,195.73	€4,540,826.78	8.64	
3	€2,192,625.79	€50,308,809.09	€52,501,434.88	€4,993,179.72	9.51	

Figure 27: Azure - Simple Architecture

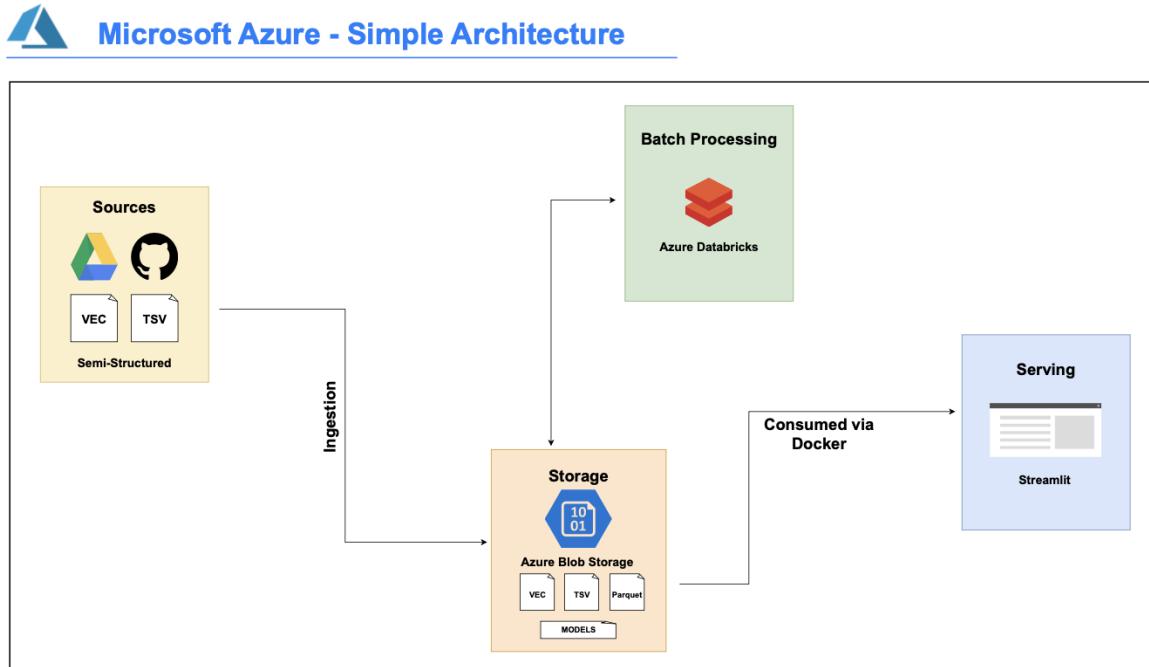


Figure 28: Azure - Low Level Architecture

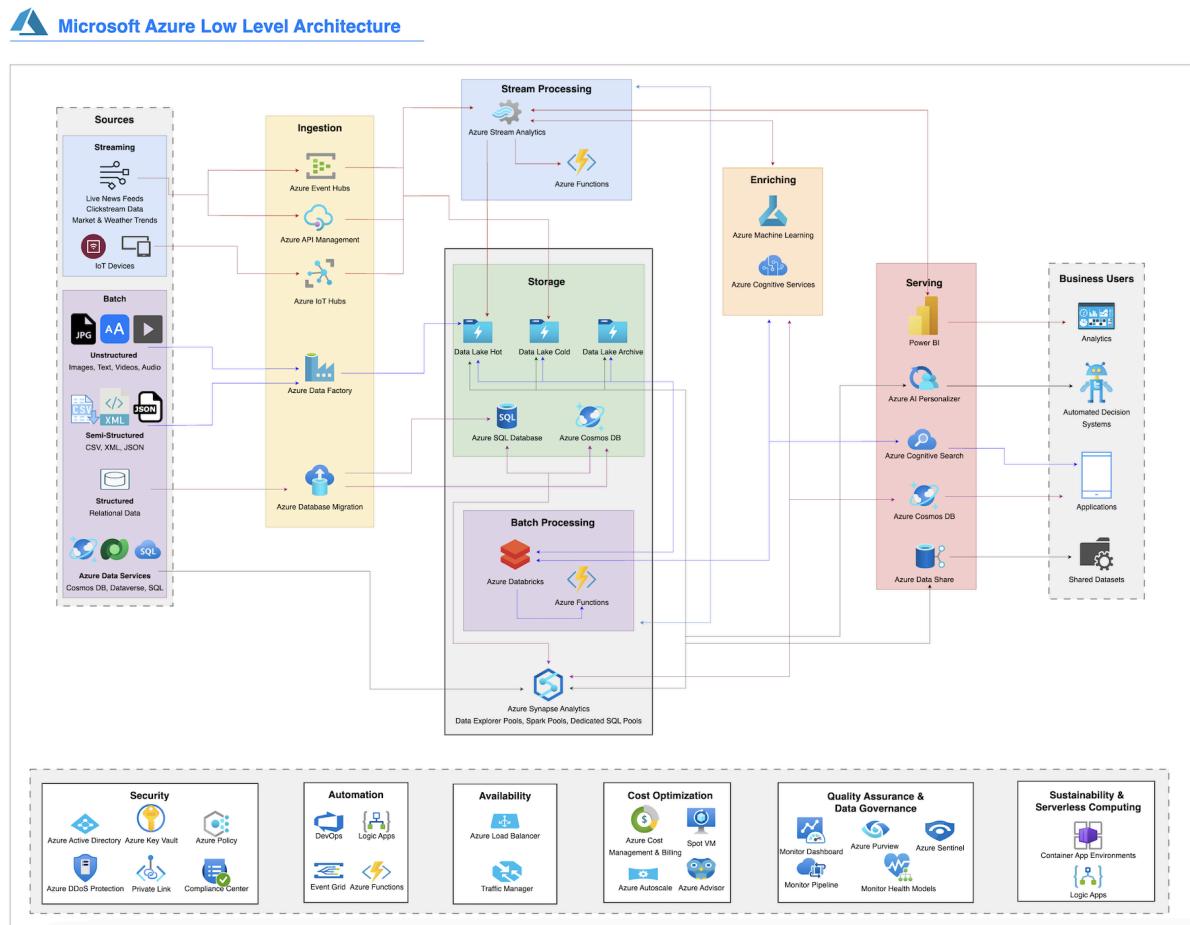


Table 15: ROI

Year	Incr. Ad Revenue	Incr. Subscription Revenue	Total Incr. Revenue	Investment	#	ROI
1	€553,693.38	€38,112,734.16	€38,666,427.54	€3,462,943.50	10.17	
2	€609,062.72	€41,924,007.58	€42,533,070.29	€3,635,431.11	10.70	
3	€730,875.26	€50,308,809.09	€51,039,684.35	€3,997,589.38	11.77	

Incr. = Incremental

Table 16: Data Ingestion Costs

Name	Cost	Comment
azure event hubs cost	50000	High-throughput event ingestion
azure iot hub cost	25000	Real-time IoT and user interaction scaling
azure api management cost	35000	Secure API handling with high request loads
azure data factory cost	50000	Increased data pipeline complexity and batch

Table 17: Storage Costs(Handling large-scale real-time and batch data)

Name	Cost	Comment
azure data lake cost	75000	Large-scale structured and unstructured storage
azure cosmos db cost	100000	Distributed real-time querying
azure sql cost	50000	Handling high-traffic relational data logs
azure synapse cost	120000	High-throughput analytics for large datasets

Table 18: Processing Costs

Name	Cost	Comment
azure databricks cost	150000	Large-scale processing for AI workloads
azure stream analytics cost	50000	Scaling real-time analytics

Table 19: AI Enrichment and Model Training Costs

Name	Cost	Comment
azure ml cost	200000	Advanced deep learning model retraining and deployment
azure cognitive services cost	100000	Scaled NLP, fake news detection, and personalization

Table 20: Deployment, Monitoring, and Compliance Costs

Name	Cost	Comment
azure devops cost	50000	Automated deployments across environments
azure logic apps cost	25000	CI/CD workflow automation
azure functions cost	30000	Serverless functions for scalable microservices
azure monitor cost	50000	High-frequency real-time monitoring
azure sentinel cost	30000	Enterprise security and intrusion detection
azure purview cost	25000	Enterprise data governance
azure key vault cost	15000	Secure secret management at scale
azure active directory cost	20000	Enterprise identity and access management
azure policy cost	10000	Organizational policy enforcement

Table 21: In-House Data Collection and Personalization Costs

Name	Cost	Comment
azure ai personalizer cost	50000	AI-based personalization scaling
azure data share cost	40000	Multi-organization data collaboration

Table 22: Security and Optimization

Name	Cost	Comment
azure container apps cost	30000	Secure AI deployment scaling
azure private links cost	25000	Secure API and data pipeline scaling
azure spot vms cost	75000	Compute cost optimization at scale