

Problem Statement

In this project I plan to try and recommend to a user a list of potential songs they will like based off of a songs musical attributes. This dataset contains about 230,000 songs from 27 different genres. I found this data set from Kaggle.com however the original data was pulled from Spotify's api open to developers

Data Characteristics

The following data attributes were used in the analysis presented here. Understanding the types of data points helped determine what cleaning was needed.

Variable Name	Description
genre	object
artist_name	object
track_name	object
track_id	object
popularity	int64
acousticness	float64
danceability	float64
duration_ms	int64
energy	float64
instrumentalness	float64
key	object
liveness	float64
loudness	float64
mode	object
speechiness	float64
tempo	float64
time_signature	object
valence	float64

Initial Data

	track name	track_id	popularity	acousticness	danceability	duration_ms	energy	instrumentalness	key	liveness	loudness	mode	speechiness	tempo	time_signature	valence
0	C'est beau de faire un show	62F10QgRfHCH2Q9uGfVfV	0	0.611	0.389	90073	0.910	0.020	Ch	0.3490	-1.828	Major	0.0203	166.969	A#4	0.814
1	Perdu d'ancee par Gas Elrhayn	62F10QgRfHCH2Q9uGfVfV	1	0.246	0.980	197373	0.737	0.000	F#	0.1010	-5.559	Minor	0.0088	174.023	A#4	0.816
2	Don't let Me be Lonely When	62F10QgRfHCH2Q9uGfVfV	2	0.902	0.882	197007	0.131	0.000	C	0.1030	-10.897	Major	0.0082	99.488	B#4	0.968
3	On my Menusier Garden Corner	62F10QgRfHCH2Q9uGfVfV	0	0.703	0.240	102427	0.326	0.000	Ch	0.0985	-12.178	Major	0.0395	171.738	A#4	0.227
4	Couverture	62u4dMRfHCH2Q9uGfVfV	4	0.950	0.331	82625	0.225	0.123	F	0.2020	-21.150	Major	0.0468	142.576	A#4	0.380

```
df.size
51: 4189050

df.shape
51: (232725, 18)
```

Data Cleaning

The majority of my cleaning at this stage was subsetting the data so that there was an equal amount of songs from each genre. This was important for when I would eventually run a clustering model.

```
df = df[df['genre'] == 'Soundtrack'].sort_values('popularity', ascending=False)[0:150]
df = df[df['genre'] == 'Indie'].sort_values('popularity', ascending=False)[0:150]
df = df[df['genre'] == 'Jazz'].sort_values('popularity', ascending=False)[0:150]
df = df[df['genre'] == 'Pop'].sort_values('popularity', ascending=False)[0:150]
df = df[df['genre'] == 'Electronic'].sort_values('popularity', ascending=False)[0:150]
df = df[df['genre'] == 'Children's Music'].sort_values('popularity', ascending=False)[0:150]
df = df[df['genre'] == 'Folk'].sort_values('popularity', ascending=False)[0:150]
df = df[df['genre'] == 'Hip-Hop'].sort_values('popularity', ascending=False)[0:150]
df = df[df['genre'] == 'Rock'].sort_values('popularity', ascending=False)[0:150]
df = df[df['genre'] == 'Alternative'].sort_values('popularity', ascending=False)[0:150]
df = df[df['genre'] == 'Classical'].sort_values('popularity', ascending=False)[0:150]
df = df[df['genre'] == 'World'].sort_values('popularity', ascending=False)[0:150]
df = df[df['genre'] == 'Blues'].sort_values('popularity', ascending=False)[0:150]
df = df[df['genre'] == 'Opera'].sort_values('popularity', ascending=False)[0:150]
df = df[df['genre'] == 'Children's Music'].sort_values('popularity', ascending=False)[0:150]
df = df[df['genre'] == 'A Cappella'].sort_values('popularity', ascending=False)[0:150]
```

#I removed comedy and a cappella from the data set, a cappella was too small and comedy was not applicable to musical song recommendations.

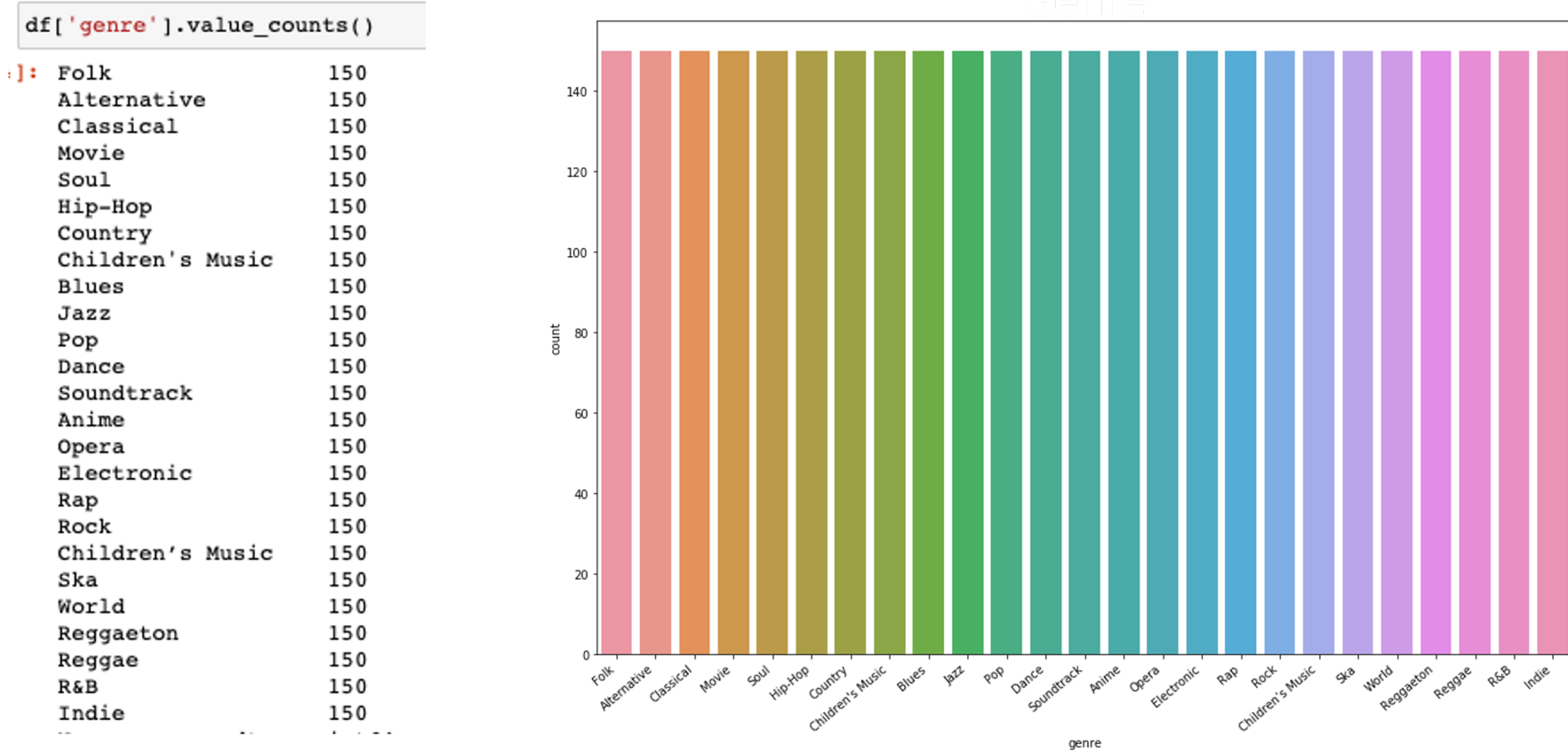
```
frames = [df1, df2, df3, df4, df5, df6, df7, df8, df9, df10, df11, df12, df13, df14, df15, df16, df17, df18, df19, df20, df21, df22, df23, df24, df25, df26, df27]
df = pd.concat(frames)
```

```
df.shape
11: 17350, 18

df.size
11: 47550
```

df['genre'].value_counts()	#Checking For NA (df.isnull().any())	# Checking For Null Values (df.isnull().sum())
11 Comedy 9461	11 genre False	11 genre 0
Soundtrack 9461	11 artist_name False	11 artist_name 0
Indie 9461	11 track_name False	11 track_name 0
Jazz 9461	11 track_id False	11 track_id 0
Pop 9386	11 popularity False	11 popularity 0
Electronic 9377	11 acousticness False	11 acousticness 0
Children's Music 9353	11 danceability False	11 danceability 0
Folk 9299	11 duration_ms False	11 duration_ms 0
Hip-Hop 9295	11 energy False	11 energy 0
Rock 9272	11 instrumentalness False	11 instrumentalness 0
Alternative 9263	11 key False	11 key 0
Classical 9256	11 liveness False	11 liveness 0
World 9094	11 loudness False	11 loudness 0
Blues 9023	11 mode False	11 mode 0
Opera 8956	11 speechiness False	11 speechiness 0
Children's Music 8927	11 tempo False	11 tempo 0
A Cappella 8874	11 time_signature False	11 time_signature 0
119 8864	11 valence False	11 valence 0
119 8864	11 dtype: bool	11 dtype: bool

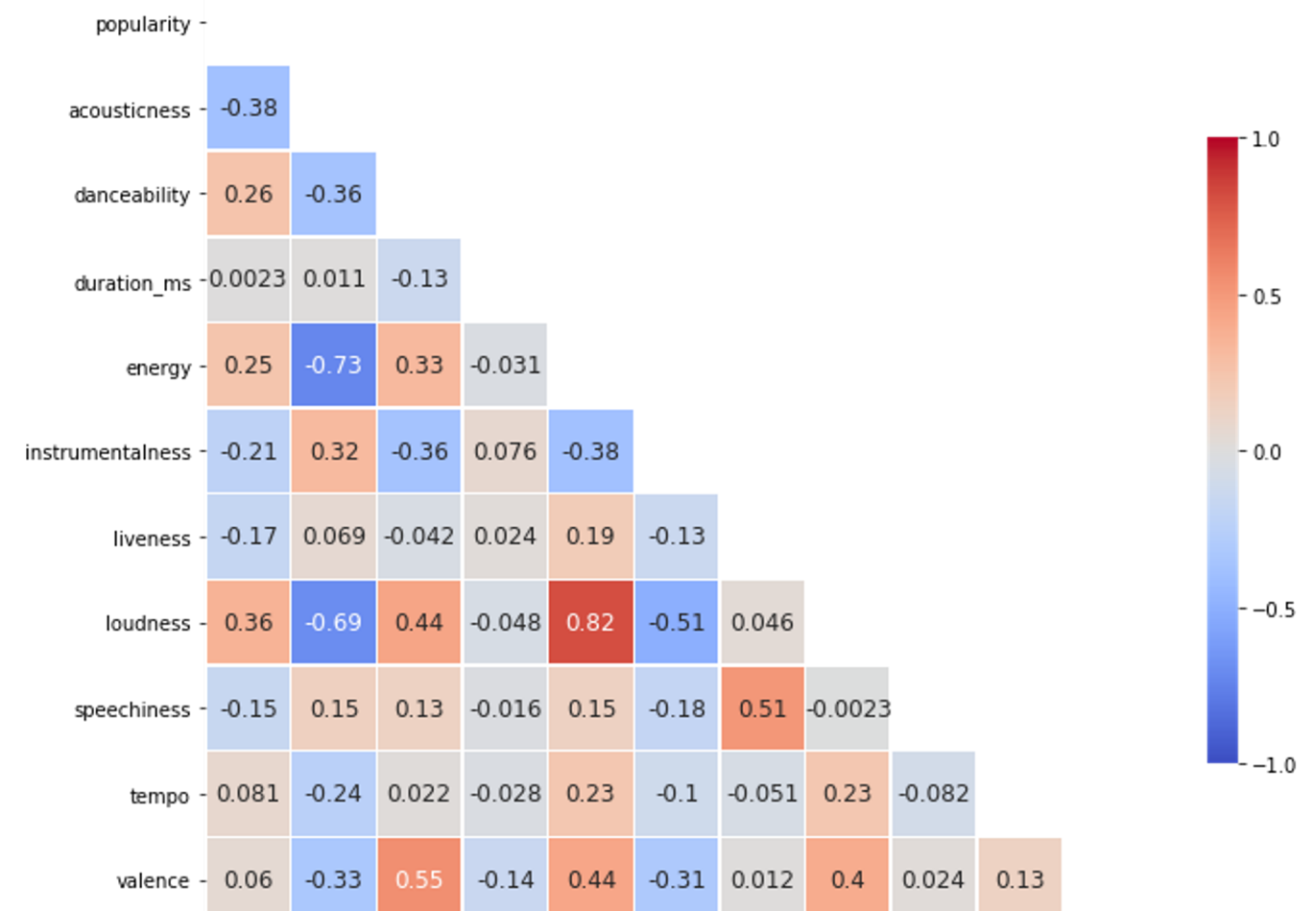
Data After Cleaning



Data Exploration

	popularity	acousticness	danceability	duration_ms	energy	instrumentalness	liveness	loudness	speechiness	tempo	valence	cluster
count	3750.000000	3750.000000	3750.000000	3750.000000	3750.000000	3750.000000	3750.000000	3750.000000	3750.000000	3750.000000	3750.000000	3750.000000
mean	68.732000	0.347904	0.675938	2.281282e+05	0.563491	0.113447	0.177132	-9.071927	0.079439	117.793168	0.467727	3.945600
std	13.718771	0.344434	0.191047	8.027557e+04	0.256471	0.279143	0.150834	6.201948	0.078269	30.891447	0.258823	3.011444
min	27.000000	0.000010	0.060900	3.155600e+04	0.001440	0.000000	0.017600	-41.808000	0.000000	35.995000	0.024000	0.000000
25%	62.000000	0.029325	0.452000	1.853402e+05	0.391000	0.000000	0.092400	-10.599500	0.026100	93.989000	0.255000	2.000000
50%	71.000000	0.207500	0.596000	2.173070e+05	0.602000	0.000008	0.116600	-7.119000	0.047300	115.074000	0.466000	3.000000
75%	78.000000	0.648750	0.720000	2.586192e+05	0.765000	0.003520	0.204000	-5.213500	0.086750	138.137500	0.671000	5.000000
max	100.000000	0.996000	0.964000	1.240600e+06	0.995000	0.992000	0.986000	1.023000	0.779000	218.184000	0.985000	10.000000

Correlation Detection



Testing Categorical Variables as Attributes for Model

```
df.columns
1: Index(['genre', 'artist_name', 'track_name', 'track_id', 'popularity', 'acousticness', 'danceability', 'duration_ms', 'energy', 'instrumentalness', 'liveness', 'loudness', 'mode', 'speechiness', 'tempo', 'time_signature', 'valence', 'key_A', 'key_B', 'key_C', 'key_D', 'key_E', 'key_F', 'key_G', 'key_H', 'key_I', 'key_J', 'key_K', 'key_L', 'key_M', 'key_N', 'key_O', 'key_P', 'key_Q', 'key_R', 'key_S', 'key_T', 'key_U', 'key_V', 'key_W', 'key_X', 'key_Y', 'key_Z', 'key_0', 'key_1', 'key_2', 'key_3', 'key_4', 'key_5', 'key_6', 'key_7', 'key_8', 'key_9', 'key_10', 'key_11', 'key_12', 'key_13', 'key_14', 'key_15', 'key_16', 'key_17', 'key_18', 'key_19', 'key_20', 'key_21', 'key_22', 'key_23', 'key_24', 'key_25', 'key_26', 'key_27', 'key_28', 'key_29', 'key_30', 'key_31', 'key_32', 'key_33', 'key_34', 'key_35', 'key_36', 'key_37', 'key_38', 'key_39', 'key_40', 'key_41', 'key_42', 'key_43', 'key_44', 'key_45', 'key_46', 'key_47', 'key_48', 'key_49', 'key_50', 'key_51', 'key_52', 'key_53', 'key_54', 'key_55', 'key_56', 'key_57', 'key_58', 'key_59', 'key_60', 'key_61', 'key_62', 'key_63', 'key_64', 'key_65', 'key_66', 'key_67', 'key_68', 'key_69', 'key_70', 'key_71', 'key_72', 'key_73', 'key_74', 'key_75', 'key_76', 'key_77', 'key_78', 'key_79', 'key_80', 'key_81', 'key_82', 'key_83', 'key_84', 'key_85', 'key_86', 'key_87', 'key_88', 'key_89', 'key_90', 'key_91', 'key_92', 'key_93', 'key_94', 'key_95', 'key_96', 'key_97', 'key_98', 'key_99', 'key_100', 'key_101', 'key_102', 'key_103', 'key_104', 'key_105', 'key_106', 'key_107', 'key_108', 'key_109', 'key_110', 'key_111', 'key_112', 'key_113', 'key_114', 'key_115', 'key_116', 'key_117', 'key_118', 'key_119', 'key_120', 'key_121', 'key_122', 'key_123', 'key_124', 'key_125', 'key_126', 'key_127', 'key_128', 'key_129', 'key_130', 'key_131', 'key_132', 'key_133', 'key_134', 'key_135', 'key_136', 'key_137', 'key_138', 'key_139', 'key_140', 'key_141', 'key_142', 'key_143', 'key_144', 'key_145', 'key_146', 'key_147', 'key_148', 'key_149', 'key_150', 'key_151', 'key_152', 'key_153', 'key_154', 'key_155', 'key_156', 'key_157', 'key_158', 'key_159', 'key_160', 'key_161', 'key_162', 'key_163', 'key_164', 'key_165', 'key_166', 'key_167', 'key_168', 'key_169', 'key_170', 'key_171', 'key_172', 'key_173', 'key_174', 'key_175', 'key_176', 'key_177', 'key_178', 'key_179', 'key_180', 'key_181', 'key_182', 'key_183', 'key_184', 'key_185', 'key_186', 'key_187', 'key_188', 'key_189', 'key_190', 'key_191', 'key_192', 'key_193', 'key_194', 'key_195', 'key_196', 'key_197', 'key_198', 'key_199', 'key_200', 'key_201', 'key_202', 'key_203', 'key_204', 'key_205', 'key_206', 'key_207', 'key_208', 'key_209', 'key_210', 'key_211', 'key_212', 'key_213', 'key_214', 'key_215', 'key_216', 'key_217', 'key_218', 'key_219', 'key_220', 'key_221', 'key_222', 'key_223', 'key_224', 'key_225', 'key_226', 'key_227', 'key_228', 'key_229', 'key_230', 'key_231', 'key_232', 'key_233', 'key_234', 'key_235', 'key_236', 'key_237', 'key_238', 'key_239', 'key_240', 'key_241', 'key_242', 'key_243', 'key_244', 'key_245', 'key_246', 'key_247', 'key_248', 'key_249', 'key_250', 'key_251', 'key_252', 'key_253', 'key_254', 'key_255', 'key_256', 'key_257', 'key_258', 'key_259', 'key_260', 'key_261', 'key_262', 'key_263', 'key_264', 'key_265', 'key_266', 'key_267', 'key_268', 'key_269', 'key_270', 'key_271', 'key_272', 'key_273', 'key_274', 'key_275', 'key_276', 'key_277', 'key_278', 'key_279', 'key_280', 'key_281', 'key_282', 'key_283', 'key_284', 'key_285', 'key_286', 'key_287', 'key_288', 'key_289', 'key_290', 'key_291', 'key_292', 'key_293', 'key_294', 'key_295', 'key_296', 'key_297', 'key_298', 'key_299', 'key_300', 'key_301', 'key_302', 'key_303', 'key_304', 'key_305', 'key_306', 'key_307', 'key_308', 'key_309', 'key_310', 'key_311', 'key_312', 'key_313', 'key_314', 'key_315', 'key_316', 'key_317', 'key_318', 'key_319', 'key_320', 'key_321', 'key_322', 'key_323', 'key_324', 'key_325', 'key_326', 'key_327', 'key_328', 'key_329', 'key_330', 'key_331', 'key_332', 'key_333', 'key_334', 'key_335', 'key_336', 'key_337', 'key_338', 'key_339', 'key_340', 'key_341', 'key_342', 'key_343', 'key_344', 'key_345', 'key_346', 'key_347', 'key_348', 'key_349', 'key_350', 'key_351', 'key_352', 'key_353', 'key_354', 'key_355', 'key_356', 'key_357', 'key_358', 'key_359', 'key_360', 'key_361', 'key_362', 'key_363', 'key_364', 'key_365', 'key_366', 'key_367', 'key_368', 'key_369', 'key_370', 'key_371', 'key_372', 'key_373', 'key_374', 'key_375', 'key_376', 'key_377', 'key_378', 'key_379', 'key_380', 'key_381', 'key_382', 'key_383', 'key_384', 'key_385', 'key_386', 'key_387', 'key_388', 'key_389', 'key_390', 'key_391', 'key_392', 'key_393', 'key_394', 'key_395', 'key_396', 'key_397', 'key_398', 'key_399', 'key_400', 'key_401', 'key_402', 'key_403', 'key_404', 'key_405', 'key_406', 'key_407', 'key_408', 'key_409', 'key_410', 'key_411', 'key_412', 'key_413', 'key_414', 'key_415', 'key_416', 'key_417', 'key_418', 'key_419', 'key_420', 'key_421', 'key_422', 'key_423', 'key_424', 'key_425', 'key_426', 'key_427', 'key_428', 'key_429', 'key_430', 'key_431', 'key_432', 'key_433', 'key_434', 'key_435', 'key_436', 'key_437', 'key_438', 'key_439', 'key_440', 'key_441', 'key_442', 'key_443', 'key_444', 'key_445', 'key_446', 'key_447', 'key_448', 'key_449', 'key_450', 'key_451', 'key_452', 'key_453', 'key_454', 'key_455', 'key_456', 'key_457', 'key_458', 'key_459', 'key_460', 'key_461', 'key_462', 'key_463', 'key_464', 'key_465', 'key_466', 'key_467', 'key_468', 'key_469', 'key_470', 'key_471', 'key_472', 'key_473', 'key_474', 'key_475', 'key_476', 'key_477', 'key_478', 'key_479', 'key_480', 'key_481', 'key_482', 'key_483', 'key_484', 'key_485', 'key_486', 'key_487', 'key_488', 'key_489', 'key_490', 'key_491', 'key_492', 'key_493', 'key_494', 'key_495', 'key_496', 'key_497', 'key_498', 'key_499', 'key_500', 'key_501', 'key_502', 'key_503', 'key_504', 'key_505', 'key_506', 'key_507', 'key_508', 'key_509', 'key_510', 'key_511', 'key_512', 'key_513', 'key_514', 'key_515', 'key_516', 'key_517', 'key_518', 'key_519', 'key_520', 'key_521', 'key_522', 'key_523', 'key_524', 'key_525', 'key_526', 'key_527', 'key_528', 'key_529', 'key_530', 'key_531', 'key_532', 'key_533', 'key_534', 'key_535', 'key_536', 'key_537', 'key_538', 'key_539', 'key_540', 'key_541', 'key_542', 'key_543', 'key_544', 'key_545', 'key_546', 'key_547', 'key_548', 'key_549', 'key_550', 'key_551', 'key_552', 'key_553', 'key_554', 'key_555', 'key_556', 'key_557', 'key_558', 'key_559', 'key_560', 'key_561', 'key_562', 'key_563', 'key_564', 'key_565', 'key_566', 'key_567', 'key_568', 'key_569', 'key_570', 'key_571', 'key_572', 'key_573', 'key_574', 'key_575', 'key_576', 'key_577', 'key_578', 'key_579', 'key_580', 'key_581', 'key_582', 'key_583', 'key_584', 'key_585', 'key_586', 'key_587', 'key_588', 'key_589', 'key_590', 'key_591', 'key_592', 'key_593', 'key_594', 'key_595', 'key_596', 'key_597', 'key_598', 'key_599', 'key_600', 'key_601', 'key_602', 'key_603', 'key_604', 'key_605', 'key_606', 'key_607', 'key_608', 'key_609', 'key_610', 'key_611', 'key_612', 'key_613', 'key_614', 'key_615', 'key_616', 'key_617', 'key_618', 'key_619', 'key_620', 'key_621', 'key_622', 'key_623', 'key_624', 'key_625', 'key_626', 'key_627', 'key_628', 'key_629', 'key_630', 'key_631', 'key_632', 'key_633', 'key_634', 'key_635', 'key_636', 'key_637', 'key_638', 'key_639', 'key_640', 'key_641', 'key_642', 'key_643', 'key_644', 'key_645', 'key_646', 'key_647', 'key_648', 'key_649', 'key_650', 'key_651', 'key_652', 'key_653', 'key_654', 'key_655', 'key_656', 'key_657', 'key_658', 'key_659', 'key_660', 'key_661', 'key_662', 'key_663', 'key_664', 'key_665', 'key_666', 'key_667', 'key_668', 'key_669', 'key_670', 'key_671', 'key_672', 'key_673', 'key_674', 'key_675', 'key_676', 'key_677', 'key_678', 'key_679', 'key_680', 'key_681', 'key_682', 'key_683', 'key_684', 'key_685', 'key_686', 'key_687', 'key_688', 'key_689', 'key_690', 'key_691', 'key_692', 'key_693', 'key_694', 'key_695', 'key_696', 'key_697', 'key_698', 'key_699', 'key_700', 'key_701', 'key_702', 'key_703', 'key_704', 'key_705', 'key_706', 'key_707', 'key_708', 'key_709', 'key_710', 'key_711', 'key_712', 'key_713', 'key_714', 'key_715', 'key_716', 'key_717', 'key_718', 'key_719', 'key_720', 'key_721', 'key_722', 'key_723', 'key_724', 'key_725', 'key_726', 'key_727', 'key_728', 'key_729', 'key_730', 'key_731', 'key_732', 'key_733', 'key_734', 'key_735', 'key_736', 'key_737', 'key_738', 'key_739', 'key_740', 'key_741', 'key_742', 'key_743', 'key_744', 'key_745', 'key_746', 'key_747', 'key_748', 'key_749', 'key_750', 'key_751', 'key_752', 'key_753', 'key_754', 'key_755', 'key_756', 'key_757', 'key_758', 'key_759', 'key_760', 'key_761', 'key_762', 'key_763', 'key_764', 'key_765', 'key_766', 'key_767', 'key_768', 'key_769', 'key_770', 'key_771', 'key_772', 'key_773', 'key_774', 'key_775', 'key_776', 'key_777', 'key_778', 'key_779', 'key_780', 'key_781', 'key_782', 'key_783', 'key_784', 'key_785', 'key_786', 'key_787', 'key_788', 'key_789', 'key_790', 'key_791', 'key_792', 'key_793', 'key_794', 'key_795', 'key_796', 'key_797', 'key_798', 'key_799', 'key_800', 'key_801', 'key_802', 'key_803', 'key_804', 'key_805', 'key_806', 'key_807', 'key_808', 'key_809', 'key_810', 'key_811', 'key_812', 'key_813', 'key_814', 'key_815', 'key_816', 'key_817', 'key_818', 'key_819', 'key_820', 'key_821', 'key_822', 'key_823', 'key_824', 'key_825', 'key_826', 'key_827', 'key_828', 'key_829', 'key_830', 'key_831', 'key_832', 'key_833', 'key_834', 'key_835', 'key_836', 'key_837', 'key_838', 'key_839', 'key_840', 'key_841', 'key_842', 'key_843', 'key_844', 'key_845', 'key_846', 'key_847', 'key_848', 'key_849', 'key_850', 'key_851', 'key_852', 'key_853', 'key_854', 'key_855', 'key_856', 'key_857', 'key_858', 'key_859', 'key_860', 'key_861', 'key_862', 'key_863', 'key_864', 'key_865', 'key_866', 'key_867', 'key_868', 'key_869', 'key_870', 'key_871', 'key_872', 'key_873', 'key_874', 'key_875', 'key_876', 'key_877', 'key_878', 'key_879', 'key_880', 'key_881', 'key_882', 'key_883', 'key_884', 'key_885', 'key_886', 'key_887', 'key_888', 'key_889', 'key_890', 'key_891', 'key_892', 'key_893', 'key_894', 'key_895', 'key_896', 'key_897', 'key_898', 'key_899', 'key_900', 'key_901', 'key_902', 'key_903', 'key_904', 'key_905', 'key_906', 'key_907', 'key_908', 'key_909', 'key_910', 'key_911', 'key_912', 'key_913', 'key_914', 'key_915', 'key_916', 'key_917', 'key_918', 'key_919', 'key_920', 'key_921', 'key_922', 'key_923', 'key_924', 'key_925', 'key_926', 'key_927', 'key_928', 'key_929', 'key_930', 'key_931', 'key_932', 'key_933', 'key_934', 'key_935', 'key_936', 'key_937', 'key_938', 'key_
```