Sentiments Analysis (Logistic Regression)

May 15, 2020

Data Analysis and Visualization

```python
[1]: # importing require libraries
     import numpy as np
     import pandas as pd
     import matplotlib.pyplot as plt
     %matplotlib inline

     import seaborn as sns
```

```python
[2]: # load dataset
     df = pd.read_csv('sentiment.csv')
     df.head()
```

```
[2]:                                                Phrase     label
     0  Spider-man is better than any summer blockbust…  Positive
     1                                   of good sense  Positive
     2  , then knock yourself out and enjoy the big sc…  Negative
     3           So could young romantics out on a date .  Positive
     4                                         humour  Positive
```

```python
[3]: # shape of the dataset
     df.shape
```

```
[3]: (2800, 2)
```

```python
[4]: # checking how many positive or negative values are there
     df.label.value_counts()
```

```
[4]: Positive    1800
     Negative    1000
     Name: label, dtype: int64
```

```python
[5]: # checking the null values
     df.isnull().sum()
```

```
[5]: Phrase    0
     label     0
     dtype: int64
```

```python
[6]: # finding the unique values from the phrase
     unique_words = list(df['Phrase'].str.split(' ', expand=True).stack().unique())

     print(unique_words)
```

```
['Spider-man', 'is', 'better', 'than', 'any', 'summer', 'blockbuster', 'we',
'had', 'to', 'endure', 'last', ',', 'and', 'hopefully', 'sets', 'the', 'tone',
'for', 'a', 'of', 'good', 'stuff', '.', 'sense', 'then', 'knock', 'yourself',
'out', 'enjoy', 'big', 'screen', 'postcard', 'that', 'self-glorified', 'Martin',
'Lawrence', 'lovefest', 'So', 'could', 'young', 'romantics', 'on', 'date',
'humour', 'You', 'bet', 'there', 'it', "'s", 'what', 'makes', 'this', 'rather',
'convoluted', 'journey', 'worth', 'taking', 'tries', 'more', 'Screenwriter',
'Chris', 'ver', 'Weil', 'directing', 'debut', 'good-natured', 'never', 'dull',
'but', 'mood', 'in', 'which', 'an', 'ominous', 'pervasive', 'unknown', 'threat',
'lurks', 'just', 'below', 'proceedings', 'adds', 'almost', 'constant',
'mindset', 'suspense', 'centers', 'midlife', 'crisis', 'hate', 'putting',
'together', 'movies', 'particular', 'value', 'or', 'merit', 'sensual', 'funny',
'Some', 'computer', 'animation', 'handsome', 'right', 'questions', 'solid',
'--', 'as', 'if', 'director', 'trying', 'dupe', 'viewer', 'into', 'all', 'Very',
'Important', 'simply', 'because', 'movie', 'ugly', 'look', 'at', 'not',
'Hollywood', 'product', 'thriller', 'once', 'too', 'often', 'war', 'One',
'year', 'most', 'weirdly', 'engaging', 'unpredictable', 'character', 'pieces',
'brilliant', 'they', 'do', "n't", 'fit', 'well', 'keep', 'interested', 'proves',
'sometimes', 'dreams', 'youth', 'should', 'remain', 'hack', 'Fire', 'nostalgic',
'twisty', 'yarn', 'will', 'them', 'guessing', 'brings', 'proper', 'conviction',
'his', 'role', '-LRB-', 'Jason', 'Bourne', '-RRB-', 'delicious', 'pulpiness',
'its', 'lurid', 'fiction', 'stylized', 'with', 'touch', 'John', 'Woo', 'bullet',
'ballet', 'you', "'re", 'entirely', 'unprepared', 'clever', 'credits', 'roll',
'ploughing', 'same', 'furrow', 'gently', 'swaying', 'back', 'forth', 'cradles',
'characters', 'veiling', 'tension', 'beneath', 'otherwise', 'tender',
'movements', '`', 'laugh', 'therapy', "'", 'whose', 'view', 'America',
'history', 'awkwardness', 'human', 'life', 'generous', 'deep', 'like', 'work',
'artist', 'who', 'tired', 'fighting', 'fights', 'weight', 'world', 'shoulders',
'playing', 'narrative', 'form', 'pair', 'fascinating', 'performances',
'Phillip', 'Noyce', 'actors', 'cinematographer', 'Christopher', 'Doyle',
'understand', 'delicate', 'forcefulness', 'Greene', 'prose', 'impossible',
'care', 'about', 'very', 'impression', 'evoke', 'memories', 'emotions', 'are',
'anything', 'drag', 'reek', 'script', 'A', 'well-made', 'humane', 'children',
'film', 'truest', 'plays', 'dramatic', 'Invigorating', 'surreal', 'resonant',
'rainbow', 'emotion', 'might', 'Vulgar', 'bilingual', 'charmer', 'can',
'honestly', 'describe', 'looking', 'sounding', 'feeling', 'no', 'other',
'recent', '``', 'Austin', 'Powers', 'Goldmember', "''", 'has', 'silly',
'entertainment', 'enough', 'laughs', 'sustain', 'interest', 'end', 'aims', 'be',
```

'uplifting', 'moving', 'occasional', 'smiles', 'surprisingly', 'achievement',
'really', 'special', 'walk', 'woods', 'show', 'us', 'time', 'courage', 'ideas',
'technical', 'proficiency', 'evil', 'monstrous', 'lunatic', 'strong', 'themes',
'familial', 'ties', 'well-intentioned', 'effort', 'still', 'burdened', 'by',
'actor', 'sneaky', 'feel', 'several', 'old', 'unintentional', 'parallels',
'inadvertently', 'humorous', 'so', 'intimate', 'psychologically', 'self-
revealing', 'passes', 'sex', 'cheap', 'hysterics', 'This', 'some', 'honest',
'insight', 'black-and-white', 'inspires', 'cliche', 'happy', 'ending',
'Swedish', 'fillm', 'modern', 'city', 'first', 'computer-generated', 'feature',
'cartoon', 'have', 'been', 'funnier', 'released', 'outtakes', 'theatrically',
'used', 'bonus', 'DVD', 'visual', 'charm', 'would', 'hard', 'sitting',
'through', 'one', 'spirals', 'downward', 'inoffensive', 'goose', 'fun', 'over-
indulgent', 'tirade', 'love', 'story', 'sanguine', 'title', 'More',
'vaudeville', 'well-constructed', 'Takes', 'clunky', 'TV-movie', 'approach',
'detailing', 'chapter', 'celebrated', 'Irish', 'playwright', 'poet', 'drinker',
'confident', 'filmmaking', 'much-needed', 'levity', 'disappointed', 'lets',
'brush', 'up', 'against', 'humanity', 'psycho', 'without', 'making', 'him',
'less', 'the-night', 'pat', 'storylines', 'precious', 'circumstances',
'beautiful', 'stars', 'success', 'wonderful', 'Madonna', 'gives', 'her', 'best',
'performance', 'since', 'Abel', 'Ferrara', 'beaten', 'pulp', 'Dangerous',
'Game', 'Eight', 'Legged', 'Freaks', 'perfectly', 'entertaining', 'diversion',
'contemplation', 'fused', 'eerie', 'atmosphere', 'deceit', 'insanely',
'violent', 'graphic', 'Starts', 'mediocre', 'thuds', 'bottom', 'pool',
'utterly', 'incompetent', 'conclusion', 'An', 'unsettling', 'memorable',
'cinematic', 'experience', 'coming', 'down', 'off', 'Miramax', 'shelves',
'after', 'couple', 'aborted', 'attempts', 'sweet', 'At', 'when',
'commercialism', 'squeezed', 'whatever', 'idealism', 'American', 'moviemaking',
'ever', 'Godfrey', 'Reggio', 'career', 'shines', 'lonely', 'beacon', 'tryingly',
'Keenly', 'observed', 'refreshingly', 'natural', 'mainstream', 'audiences',
'rarely', 'seen', 'indication', 'Got', 'depress', 'itself', 'Superbly',
'Filmmaker', 'Tian', 'Zhuangzhuang', 'triumphantly', 'returns', 'visually',
'masterful', 'quiet', 'power', 'captures', 'innocence', 'budding', 'demons',
'within', 'wallflower', 'created', 'monster', 'did', 'know', 'how', 'handle',
'If', 'horror', 'primary', 'goal', 'frighten', 'disturb', 'They', 'works',
'spectacularly', '…', 'shiver-inducing', 'nerve-rattling', 'ride', 'give',
'accurate', 'thought', 'right-thinking', 'films', 'revel', 'splendor', 'Feels',
'Even', 'fans', 'Ismail', 'Merchant', 'I', 'suspect', 'indie', 'powerful',
'people', 'fully', 'does', 'leave', 'much', 'ham-fisted', 'introspective',
'independent', 'thoughtful', 'art', 'ethics', 'cost', 'moral', 'compromise',
'B', 'Perhaps', 'call', 'refreshing', 'endless', 'assault', 'embarrassingly',
'jokes', 'Reeks', 'rot', 'home', 'French', 'shocker', 'usual', 'bad', 'boy',
'weirdo', 'team', 'offbeat', 'humor', 'amusing', 'warmest', 'relic', 'from',
'bygone', 'era', 'convolutions', 'plausible', 'bizarre', 'Although', 'fairly',
'involving', 'far', 'goes', 'having', 'fresh', 'say', 'growing', 'Catholic',
'uncluttered', 'gem', 'leaps', 'over', 'national', 'boundaries', 'celebrates',
'universal', 'nature', 'waxes', 'poetic', 'our', 'taste', 'amazing',
'slapstick', 'instrument', 'different', 'own', 'yet', 'instantly',
'recognizable', 'heroes', 'try', 'avoid', 'succeed', 'few', 'cool', 'seems',

'aware', 'coolness', 'little', 'recommend', 'Snow', 'Dogs', 'Waking', 'Up',
'Reno', 'case', 'letting', 'sleeping', 'dogs', 'lie', 'The', 'rehashes',
'capped', 'pointless', 'extremes', 'hellish', 'conditions', 'feels', 'pretty',
'execution', 'pomposity', 'solving', 'problem', 'distract', 'solution',
'another', 'poorly', 'delivered', 'smug', 'self-satisfaction', 'usually',
'associated', 'private', 'schools', 'alone', 'force', 'scratch', 'hole', 'your',
'head', 'experimental', 'storytelling', 'losing', 'stooping', 'base',
'melodrama', 'decent', 'material', 'music', 'lot', 'richer', 'ones', 'action',
'screenwriters', 'come', 'their', 'loosely-connected', 'string', 'mythic',
'constructs', 'relays', 'points', 'lectures', 'confrontations', 'beyond',
'comprehension', 'even', 'were', 'Moonlight', 'Mile', 'judgment', 'damned',
'Hartley', 'emphasizes', 'well-wrought', 'omits', 'needless', 'chase', 'scenes',
'swordfights', 'revenge', 'unfolds', 'all-enveloping', 'Provides', 'kind',
'need', 'comedies', 'spiritual', 'riveting', 'set', 'comedy', 'equivalent',
'song', 'finally', 'found', 'vent', 'Importance', 'Being', 'Earnest', 'thick',
'wit', 'reading', 'Bartlett', 'Familiar', 'Quotations', 'gritty', 'major',
'weird', 'Masterpiece', 'Theater', 'sketch', 'neither', 'concentrating',
'elements', 'revealing', 'alienation', 'among', 'culture', 'sadly', 'hostile',
'odds', 'recklessness', 'retaliation', 'When', 'rich', 'luscious', 'nuances',
'empty', 'intrigue', 'betrayal', 'murder', 'light', 'Errol', 'Morris',
'focusing', 'eccentricity', 'failing', 'ultimately', 'make', 'something',
'bigger', 'scrapbook', 'oddballs', 'attract', 'upscale', 'hungry', 'quality',
'corny', 'examination', 'affecting', 'portrait', 'screwed-up', 'man', 'dared',
'mess', 'Written', 'flatly', 'Mr.', 'Polanski', 'element', 'here', 'zingers',
'barely', 'defensible', 'sexual', 'violence', 'Inventive', 'intoxicatingly',
'sexy', 'self-indulgent', 'maddening', 'example', 'lush', 'predictable',
'reassuring', 'manner', 'beautifully', 'sung', 'holiday', 'carol', 'vapid',
'exercise', 'Best', 'delivers', 'perfect', 'get', 'breezy', 'blend', 'banality',
'big-screen', 'remakes', 'Avengers', 'Wild', 'West', 'crafted', 'wisecracking',
'Mystery', 'Science', '3000', 'guys', 'keeps', 'IMAX', 'short', 'X', 'blab',
'what-if', 'photographed', 'staged', 'Mendes', 'series', 'vulgar',
'forgettably', 'dark', 'real', 'reason', 'made', 'Mamet', 'House', 'Games',
'fall', 'Heist', 'prevention', 'place', 'blame', 'baseball', 'macabre', 'where',
'religious', 'civic', 'virtues', 'hold', 'society', 'tatters', 'Both', 'lead',
'Oscar-size', 'actually', 'filmmakers', 'want', 'nothing', 'else', 'way',
'Whenever', 'threatens', 'bogged', 'earnest', 'dramaturgy', 'fairy', 'tale',
'comes', 'renowned', 'Indian', 'allows', 'Americans', 'preposterous', 'moments',
'seeking', 'Like', 'sizzle', 'news', 'exotic', 'creatures', 'overboard',
'loony', 'melodramatic', 'denouement', 'high', 'school', 'swimming',
'substitutes', 'bathtub', 'evanescent', 'seamless', 'sumptuous', 'stream',
'consciousness', 'only', 'quest', 'president', 'single-handedly', 'turned',
'plane', 'full', 'hard-bitten', 'cynical', 'journalists', 'was', 'essentially',
'campaign', 'extended', 'publicity', 'department', 'graced', 'Poor', 'editing',
'bluescreen', 'ultra-cheesy', 'dialogue', 'complex', 'those', 'crass',
'contrived', 'sequels', 'fails', 'second-guess', 'affection', 'original',
'position', 'Thanks', 'largely', 'emotionally', 'least', 'Wiseman', 'fatal',
'error', 'flash', 'rock', 'videos', 'rip-roaring', 're-creations', 'famous',
'gratuitous', 'distractions', 'chuckles', 'gag', 'positively', 'thrilling',

'combination', 'ethnography', 'good-bad', 'Eerily', 'moment', 'high-concept',
'candy-coat', 'spirituality', 'definite', 'room', 'improvement', 'Robert',
'Harmon', 'less-is-more', 'bump-in', '-', 'chills', 'greatest', 'triumph',
'keeping', 'creepy', 'crawlies', 'hidden', 'shadows', 'hews', 'carries',
'effortlessly', 'darkness', 'wondering', 'lives', 'offers', 'opportunities',
'Cinema', 'Paradiso', 'find', 'new', 'interesting', "'d", 'probably', 'turn',
'convinced', 'already', 'employ', 'kids', 'owe', 'favors', 'parents',
'companion', 'Its', 'generic', 'villains', 'lack', 'accents', 'cinema',
'wondrously', 'gifted', 'artists', 'Arthur', 'Schnitzler', 'Reigen', 'Sy',
'open-faced', 'smiling', 'madmen', 'difficult', 'worthy', 'airless', 'shell',
'games', 'latest', 'eccentric', 'super-wealthy', 'megalomaniac', 'bent',
'domination', 'destruction', 'romantic', 'Proves', 'formulas', 'messing',
'serves', 'workable', 'primer', 'region', 'Manages', 'please', 'intended',
'audience', 'placing', 'coma-like', 'state', 'occurs', '7', 'times', 'during',
'Windtalkers', 'serious-minded', 'modest', 'winning', 'misses', 'prisoner',
'Smoke', 'Signals', 'also', 'imbued', 'paralyzed', 'stoner', 'midnight',
'flick', 'sci-fi', 'deconstruction', 'cliches', 'considerable', 'dash',
'running', 'around', 'screaming', 'death', 'style', 'sincere', 'veers', 'comic',
'course', 'It', 'gets', 'doubt', 'asks', 'country', 'serious', 'ways',
'extravagant', 'chance', 'distort', 'perspective', 'throw', 'path',
'incoherence', 'despite', 'impressed', 'upon', 'injects', 'freshness', 'acting-
workshop', 'exercises', 'Saddam', 'Hussein', 'In', 'cut', 'rhapsodizes', 'add',
'bleak', 'Trouble', '25', 'minutes', 'gratitude', 'basic', 'comedic', 'likes',
'franchise', 'years', 'small', 'easily', 'overshadowed', 'predictability',
'esoteric', 'musings', 'philosophy', 'emotional', 'connections', 'absorbing',
'Scarcely', 'expense', 'paid', 'pay', 'see', 'shows', 'slice', 'production',
'values', 'Malcolm', 'D.', 'Lee', 'writer', 'Ridley', 'Downright',
'transparent', 'rewrite', 'designed', 'garner', 'cooler', 'PG-13', 'rating',
'welcome', 'relief', 'stop', 'mom', 'amiable', 'degree', 'randomness',
'achieved', 'lottery', 'drawing', 'plotting', 'details', 'promenade', 'clad',
'bodies', 'Myrtle', 'Beach', 'S.C.', 'adrenaline', 'jolt', 'sudden', 'lunch',
'rush', 'diner', 'unfocused', 'hero', 'open', 'devolves', 'ends', 'satire',
'mind', 'adolescent', 'demanding', 'regular', 'shocks', 'bouts', 'Mostly',
'Martha', 'mostly', 'unsurprising', 'And', 'appreciate', 'one-sided', 'theme',
'predecessors', 'sensational', 'true-crime', 'theatres', 'many', 'things',
'delectable', 'everything', 'expect', 'number', 'tumbleweeds', 'blowing',
'terms', 'bites', 'chew', 'linking', 'massacre', 'Armenians', '1915',
'relationships', 'present', 'rut', 'life-changing', 'encounters', 'bagatelle',
'nearly', 'epic', 'proportions', 'Violent', 'superior', 'stale', 'uninspired',
'toxic', 'intoxicating', 'may', 'believe', 'puts', 'battle', 'wills',
'revelatory', 'narcissistic', 'morality', 'While', 'About', 'Boy', 'provide',
'enjoyable', '100', 'theater', 'Swimfan', 'Fatal', 'Attraction', 'brilliantly',
'played', 'deeply', 'highest', 'wears', 'Mark', 'Pellington', 'pop', 'kooky',
'overeager', 'spooky', 'subtly', 'myth', 'Mediocre', "'m", 'ready', 'go',
'U.N.', 'ask', 'permission', 'preemptive', 'strike', 'occasionally',
'horrifying', 'inspiring', 'expected', 'grace', 'Miyazaki', 'storytellers',
'achieving', 'cliched', 'perverse', 'escapism', 'Charming', 'apart',
'reporting', 'flickering', 'reminders', 'bind', 'Mood', 'Beautifully', 'miss',

'joyous', 'unless', 'considers', 'source', 'hilarity', 'sure', 'filmmaker',
'disagree', 'point', 'chronicle', 'smart', 'provocative', 'drama', ':',
'mention', 'company', 'overexposed', 'waste', 'gorgeously', 'atmospheric',
'meditation', 'Beating', 'sad', 'professional', 'conventional', 'comedy-drama',
'post-camp', 'old-time', 'Run', 'Do', 'judge', 'soon', 'Bloody', 'grieving',
'process', 'transcends', 'boy-meets-girl', 'posturing', 'typical', 'stories',
'outnumber', 'hits', 'three-to-one', 'appropriate', 'compelling',
'increasingly', 'diverse', 'seeing', 'great', 'deal', 'though', 'Sillier',
'cuter', 'shorter', 'remember', 'bit', 'heavy', 'handed', 'message',
'recommended', 'viewing', 'acting', 'Spielberg', 'realization', 'near-future',
'Horrendously', 'amateurish', 'plainly', 'incomprehensible', 'recognized',
'mildly', 'Changing', 'Lanes', 'mystical', 'Robinson', 'web', 'matches', 'page-
turning', 'frenzy', 'Clancy', 'creates', 'heavy-handed', 'screenplay', 'Too',
'clumsy', 'key', 'splash', 'absolute', 'joy', 'poetry', 'common', 'through-
line', 'pinnacle', 'going', 'house', 'party', 'watching', 'host', 'defend',
'himself', 'frothing', 'ex-girlfriend', 'sophomoric', 'romp', 'happen', 'Has',
'rare', 'capability', 'soothe', 'break', 'heart', 'single', 'stroke', 'tragedy',
'record', 'tenacious', 'fighter', 'victim', 'Well-done', 'supernatural',
'catch', 'lacks', 'quick', 'Steven', 'Schindler', 'List', 'Williams',
'developments', 'processed', '60', 'intellectual', 'lector', 'auteur',
'injuries', 'Parents', 'beware', ';', 'hopeful', 'maintaining', 'appearance',
'clinical', 'objectivity', 'start', 'speeds', 'Burr', 'Steers', 'knows',
'territory', 'particularly', 'Witty', 'Those', 'saved', 'aid', 'teeth', 'hurt',
'mawkish', 'self-parody', 'hardly', 'exist', 'responsible', 'generates', 'plot',
'What', 'begins', 'evolves', 'Imax', 'chimps', 'lots', 'blown', 'size',
'latently', 'gay', 'scattered', 'bogus', 'There', 'grenade', 'rooted',
'undergoing', 'cares', '?', 'Judith', 'Zaza', 'bedroom', 'sequence', 'bulk',
'wrong', 'heartfelt', 'middle', 'dedication', 'Aggressive', 'self-
glorification', 'manipulative', 'whitewash', 'consistently', 'unimaginative',
'sophisticated', 'literate', 'such', 'pictures', 'standard', 'procedure',
'prince', 'fellow', 'survival', 'awkwardly', 'magic', 'realism', 'meaty',
'face', 'thing', 'tackled', 'subject', 'drawn', 'while', 'peppering', 'pages',
'Analyze', 'That', 'popcorn', 'Every', 'Day', 'plodding', 'Vincent', 'Gallo',
'finish', 'meaning', 'phrase', 'rambling', 'living', 'mug', 'shots',
'impeccable', 'pedigree', 'mongrel', 'pep', 'Uneven', 'nice', 'prove', 'potent',
'hell-jaunt', 'purists', 'Tolstoy', 'groupies', 'tricky', 'satisfying', 'sharp',
'homage', 'Them', 'Tarantula', 'low', '10', 'compressed', 'Tells', 'terrific',
'10th-grade', 'learning', 'tool', 'smartly', 'fact', 'Disney', 'scrape', 'me',
'successful', 'dragons', 'Reign', 'Shainberg', 'adaptation', 'gander', 'amuses',
'none', 'amounts', 'Sleek', 'arty', 'daring', 'verve', 'outpaces',
'contemporaries', 'weak', 'wounds', 'inspired', 'indecipherable',
'complications', 'he', 'revealed', 'before', 'abandoned', 'consoled',
'depiction', 'depression', 'non-techies', 'Though', 'rent', 'parable',
'inexorably', 'rote', 'sentimentality', 'Hampered', "'ve", 'feature-length',
'antics', 'pleasure', 'cracker', 'barrel', 'reaction', 'subjects', 'furiously',
'Genuinely', 'unnerving', 'mesmerizing', '', 'disguised', 'images', 'attracts',
'unnamed', 'substitutable', 'forces', 'serve', 'terror', "'ll", 'admit', 'my',
'complaint', 'Amid', 'shock', 'curiosity', 'factors', 'actress', 'idealistic',

'kid', 'chooses', 'champion', 'cause', 'pretentiousness', 'Damon', 'Wilde',
'surprising', 'lazy', 'budget', 'B-movie', 'thrillers', '1950s', "'60s",
'Campanella', 'directed', 'expedience', 'roiling', 'documentary', 'Sunday',
'flair', 'slight', 'poem', 'eyes', 'put', 'hairs', 'chest', 'told', 'liking',
'Ice', 'Age', 'given', 'day', 'comprehensible', 'Dummies', 'guide', 'Daughter',
'frame', 'Intriguing', 'downright', 'completely', 'emphasis', 'James', 'Eric',
'Horton', 'Peter', "O'Fallon", 'stirring', 'dog', 'Winds', 'being', 'both',
'surge', 'swirling', 'rapids', 'leap', 'rouses', 'mainland', 'setting',
'Transforms', 'Shakespeare', 'deepest', 'tragedies', 'proud', 'Ki-Deok',
'After', 'woman', 'run-of-the-mill', 'ruthless', 'social', 'order', 'governs',
'college', 'cliques', 'flame-like', 'trembling', 'progression', 'two',
'skittish', 'New', 'York', 'middle-agers', 'stumble', 'relationship',
'struggle', 'fears', 'foibles', 'Mary', 'Gaitskill', 'harrowing', 'improved',
'destination', 'mixed', 'results', 'escapades', 'demonstrating', 'adage', 'wo',
'safe', 'Plays', 'vignettes', 'clips', 'Pow', 'Swimming', 'radical', 'Just',
'everyone', 'involved', 'coasting', 'But', 'followed', 'runaway', 'Full',
'Monty', 'inept', 'Jones', 'Evokes', 'David', 'raunch', 'favor', 'gags', 'rely',
'strength', 'cleverness', 'Report', 'card', 'Does', 'live', 'exalted',
'tagline', 'sidekicks', 'Devos', 'Oedekerk', 'childhood', 'dream', 'martial-
arts', 'lose', 'above', 'Shakespearean', 'Less', 'efforts', 'toward', 'closure',
'highlight', 'Spalding', 'Gray', 'teen', 'gross-out', 'Love', 'Hong', 'Kong',
'From', 'Danang', 'reveals', 'scientific', 'law', 'discerned', 'producers',
'heed', 'Hatfield', 'Hicks', 'oddest', 'couples', 'becomes', 'study', 'gambles',
'publishing', 'offering', 'exists', 'political', 'ramifications', 'glacial',
'pacing', 'fan', 'fantasia', 'Monte', 'Cristo', 'war-ravaged', 'land',
'unlikely', 'Sarah', 'Harrison', 'keen', 'insights', 'parapsychological',
'phenomena', 'soulful', 'Jolting', 'under', 'skin', 'freakish', 'seem', 'saw',
'TV', '300', 'Russian', 'mighty', 'Ivan', 'swear', 'rest', 'game', 'hooked',
'As', 'Storytelling', 'version', 'Quiet', 'creature-feature', 'Shyamalan',
'boilerplate', 'punch', 'lines', 'unfortunately', 'double-cross', 'Swinging',
'hobby', 'Nothing', 'various', 'eventually', 'All', 'arch', 'cold-blooded',
'Claude', 'Chabrol', 'camera', 'Entertains', 'Kim', 'tenderness', 'takes',
'generate', 'heat', 'juicy', 'soap', 'opera', 'Cho', 'creating', 'navigates',
'fast', 'fade', '!', 'blaxploitation', 'spoof', 'downplays', 'opposed',
'extent', 'outrageousness', 'fest', 'Using', 'stock', 'scruffy', 'slip',
'convincing', 'dizzily', 'gorgeous', 'nor', 'Narratively', 'drive-by',
'tentative', 'familiar', 'anti-feminist', 'equation', '=', 'misery', 'ambition',
'whole', 'staring', 'Kendall', 'structure', 'Offers', 'Initial', 'strangeness',
'Q', 'Quirky', 'Now', 'former', 'Gong', 'Show', 'addict', 'enthusiasm', 'Not',
'scares', 'cross', 'chemicals', 'bunch', 'totally', 'unexpected', 'directions',
'Green', 'arriving', 'core', 'Tartakovsky', 'powers', 'memory', 'monsterous',
'penance', 'cowering', 'begging', 'feet', 'Giannini', 'Narc', 'strikes',
'defiantly', 'retro', 'chord', 'willingness', 'experiment', 'explosions',
'early', 'trusted', 'sort', 'toothless', 'book', 'Favor', 'document', 'Comes',
'across', 'Highly', 'Contrived', 'depressing', 'ruthlessly', 'pained',
'depraved', 'wound', 'Scherfig', 'five', 'writers', 'He', 'Fresnadillo',
'Videodrome', '100-minute', 'Something', 'co-writer', 'Ed', 'Decter', 'senses',
'mean-spirited', 'Before', 'broken', 'outweighs', 'loss', 'witness', 'visible',

```
'His', 'preach', 'exclusively', 'converted', 'Writer\\/director', 'Kung',
'Seeing', 'momentum', 'remains', 'undeterminable', 'roughshod', 'sensibilities',
'makers']
```

[7]:
```python
# count the unique word from phrase
unique_word_count = df.Phrase.str.split(' ', expand=True).stack().value_counts()

print(unique_word_count)
```

```
the          995
,            817
a            740
and          702
of           701
             ...
generate       1
Favor          1
!              1
Just           1
primary        1
Length: 1979, dtype: int64
```

Here, the must frequent word is `the` and total number of unique words is 1979.

[8]:
```python
# counting the workds in each row
df['WordCount'] = df['Phrase'].str.split().str.len()

df.head()
```

[8]:

	Phrase	label	WordCount
0	Spider-man is better than any summer blockbust…	Positive	27
1	of good sense	Positive	3
2	, then knock yourself out and enjoy the big sc…	Negative	19
3	So could young romantics out on a date .	Positive	9
4	humour	Positive	1

[9]:
```python
# checking if there is empyt Phrase or not
df[df['WordCount'] == 0]
```

[9]:

	Phrase	label	WordCount
981		Negative	0

[10]:
```python
# dropping the index which has no phrase
df.drop(df[df['WordCount'] == 0].index, inplace=True)

df.shape
```
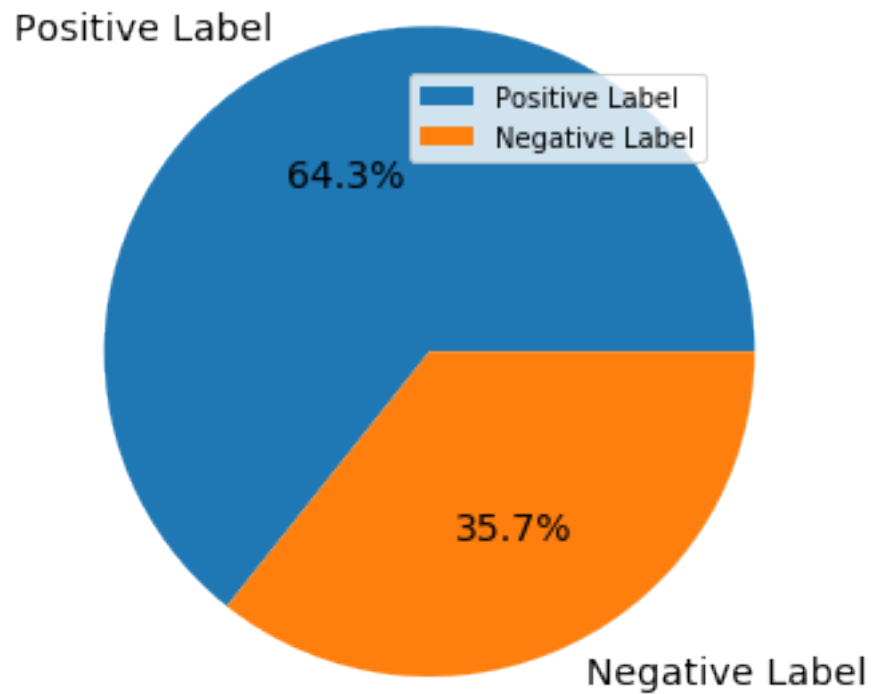
[10]: (2799, 3)

```
[11]: # plotting the pie chart
      pos_total = len(df[df['label'] == 'Positive'])
      neg_total = len(df[df['label'] == 'Negative'])

      textprops = {'fontsize': 14}
      plt.pie(x=[pos_total, neg_total], labels=['Positive Label', 'Negative Label'],␣
       ↪textprops=textprops, autopct='%0.1f%%', radius=1.4)

      plt.legend()
      plt.show()
```
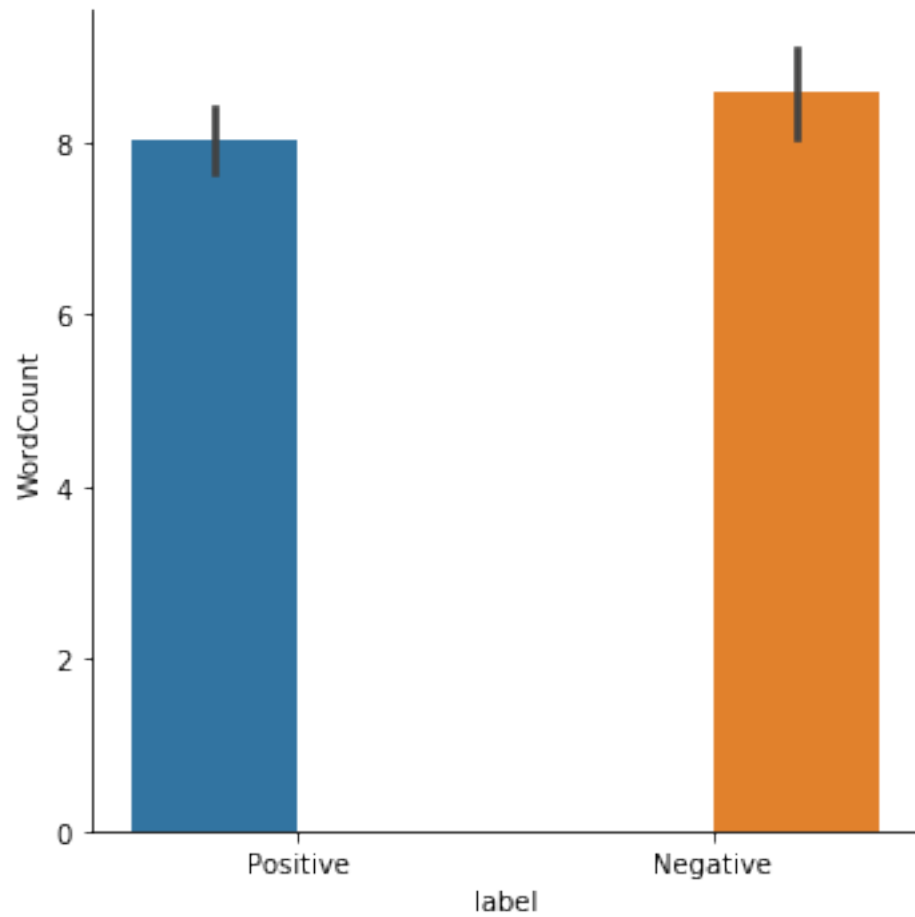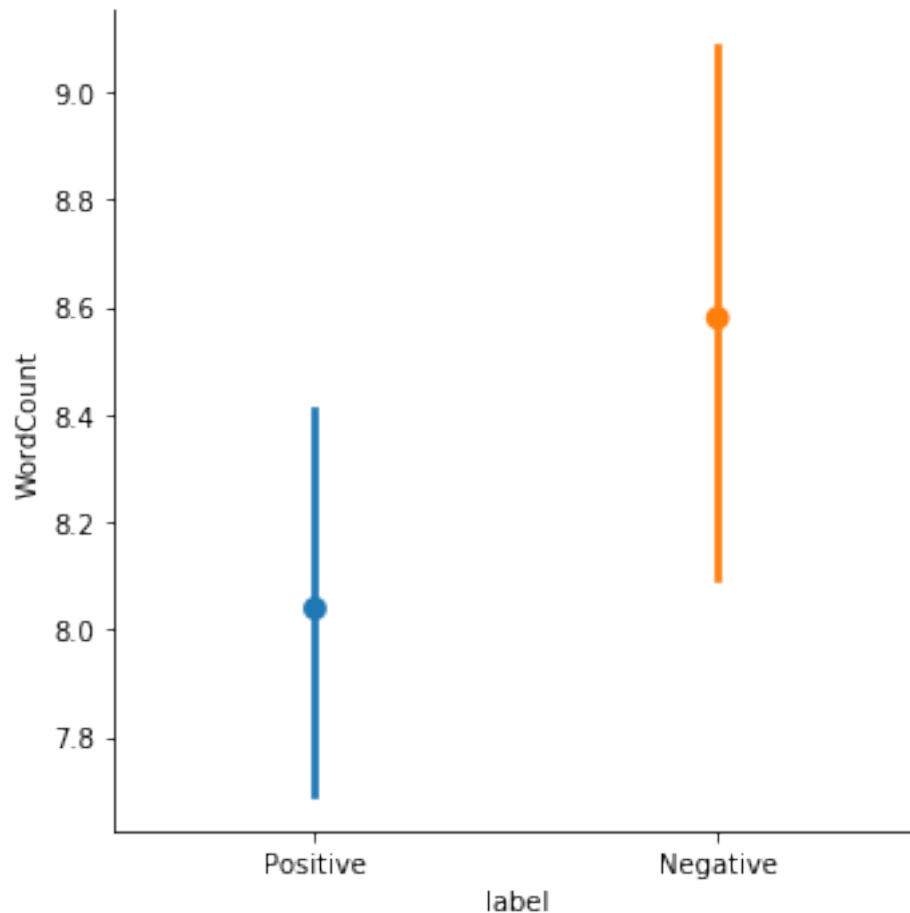


```
[12]: # average wordcount par plot

      sns.catplot(x='label', y='WordCount', data=df, kind='bar', hue='label')
      plt.show()
```

```
[13]: # plotting the categorical data

      sns.catplot(x='label', y='WordCount', data=df, kind='point', hue='label')
      plt.show()
```

The above plots shows that the `Negative` label has slightly more word count than the `Positive` word count in general.

Feature Extraction

Seperating the train, test & validation dataset.

```python
# seperating train, test and validate set
total_len = len(df)
train_len = int(0.6 * total_len)
val_len = int(0.2 * total_len)

print('Total Length: ', total_len)
print('Train Length: ', train_len)
print('Validation Length: ', val_len)
print('Test Length: ', val_len)
```

```
Total Length:  2799
Train Length:  1679
Validation Length:  559
```

Test Length: 559

```python
[19]: # splitting dataset into train, test & val datset
      train_df = df[: train_len]
      val_df = df[train_len: train_len + val_len]
      test_df = df[train_len + val_len: ]

      print('train_df len: ', len(train_df))
      print('val_df len: ', len(val_df))
      print('test_df len: ', len(test_df))
```

```
train_df len:   1679
val_df len:   559
test_df len:   561
```

Extracting unique vocal from the training dataset.

```python
[22]: vocab = list()

      for i in train_df['Phrase'].tolist():
          vocab.extend(i.split(' '))

      unique_vocab = list(set(vocab))

      print('Vocab length: ', len(vocab))
      print('Unique vocab length: ', len(unique_vocab))
```

```
Vocab length:   13853
Unique vocab length:   1843
```

```python
[23]: def feature_extraction(unique_vocab, dataset):

          # vectorizing the phrase
          feature = []
          label = []

          for index, row in dataset.iterrows():
              row_words = row['Phrase'].split(' ')

              f = []
              for v in unique_vocab:
                  if v in row_words:
                      f.append(1)
                  else:
                      f.append(0)

              if row['label'] == 'Positive':
                  label.append(1)
```

```
        elif row['label'] == 'Negative':
            label.append(0)

        feature.append(f)

    # converting into numpy array
    feature = np.array(feature)
    label = np.array(label)

    # changing label into 2d array
    label = label.reshape(-1, 1)
    print('Feature shape: ', feature.shape)

    return feature, label
```

[24]:
```
# creating X_train, y_train of feature and label from a dataset
X_train, y_train = feature_extraction(unique_vocab, train_df)
X_val, y_val = feature_extraction(unique_vocab, val_df)
X_test, y_test = feature_extraction(unique_vocab, test_df)
```

```
Feature shape:  (1679, 1843)
Feature shape:  (559, 1843)
Feature shape:  (561, 1843)
```

[25]:
```
# Adding bias to x data
X_train = np.insert(X_train, 0, values=1, axis=1)
X_val = np.insert(X_val, 0, values=1, axis=1)
X_test = np.insert(X_test, 0, values=1, axis=1)
```

Here, all our data is 0 and 1. so, we dnt have to do any normalization here.

Grid Search

[27]:
```
np.random.seed(42)

params = {
    'num_of_iteration': 1000,
    'learning_rate': 0.001
}

grid_params = {
    'num_of_iteration': [1000, 1500, 2000, 2500],
    'learning_rate': [0.001, 0.01, 0.1, 0.6]
}
```

Binary Cross Entropy Loss and Sigmoid function

```
[28]:  # bcel
       def binary_cross_entropy_loss(h, y):
           m = len(h)

           return -np.average(y * np.log(h) + (1 - y) * np.log(1 - h))



       # sigmoid function
       def sigmoid(x):
           return 1 / (1 + np.exp(-x))
```

Training Model

```
[29]:  def train_model(X_train, y_train, X_val, y_val, params):
           train_error = []
           w = np.random.rand(X_train.shape[1]).reshape(-1, 1)

           for i in range(0, params['num_of_iterations']):
               h_train = sigmoid(np.matmul(X_train, w))
               train_loss = binary_cross_entropy_loss(h_train, y_train)
               train_error.append(train_loss)

               gradient_descent = (np.matmul(np.transpose(X_train), (h_train -␣
       ↪y_train))) / X_train.shape[0]
               w = w - params['learning_rate'] * gradient_descent


           h_val = sigmoid(np.matmul(X_val, w))
           val_loss = binary_cross_entropy_loss(h_val, y_val)
           print(params, val_loss)

           return w, train_error
```

Parameter with val loss

```
[31]:  # grid search
       import itertools

       grid = list(itertools.product(grid_params['num_of_iteration'],␣
        ↪grid_params['learning_rate']))


       for g in grid:
           p = {
               'num_of_iterations': g[0],
               'learning_rate': g[1]
           }
```

```
    train_model(X_train, y_train, X_val, y_val, p)
```

```
{'num_of_iterations': 1000, 'learning_rate': 0.001} 1.4593497866218992
{'num_of_iterations': 1000, 'learning_rate': 0.01} 0.8916758730910846
{'num_of_iterations': 1000, 'learning_rate': 0.1} 0.5552828426718963
{'num_of_iterations': 1000, 'learning_rate': 0.6} 0.4058936846842115
{'num_of_iterations': 1500, 'learning_rate': 0.001} 1.3609936822386934
{'num_of_iterations': 1500, 'learning_rate': 0.01} 0.8554623226830117
{'num_of_iterations': 1500, 'learning_rate': 0.1} 0.5320832129012135
{'num_of_iterations': 1500, 'learning_rate': 0.6} 0.3846666832920364
{'num_of_iterations': 2000, 'learning_rate': 0.001} 1.2875999923591088
{'num_of_iterations': 2000, 'learning_rate': 0.01} 0.7895518938559557
{'num_of_iterations': 2000, 'learning_rate': 0.1} 0.514189997981818
{'num_of_iterations': 2000, 'learning_rate': 0.6} 0.3653870117713219
{'num_of_iterations': 2500, 'learning_rate': 0.001} 1.264794362599046
{'num_of_iterations': 2500, 'learning_rate': 0.01} 0.7482513782130533
{'num_of_iterations': 2500, 'learning_rate': 0.1} 0.4802926312744272
{'num_of_iterations': 2500, 'learning_rate': 0.6} 0.3448030512603277
```

```python
[33]: final_params = {
          'num_of_iterations': 2500,
          'learning_rate': 0.6
      }

      final_weights, train_loss = train_model(X_train, y_train, X_val, y_val,␣
        ↪final_params)
```

```
{'num_of_iterations': 2500, 'learning_rate': 0.6} 0.34336854752113327
```
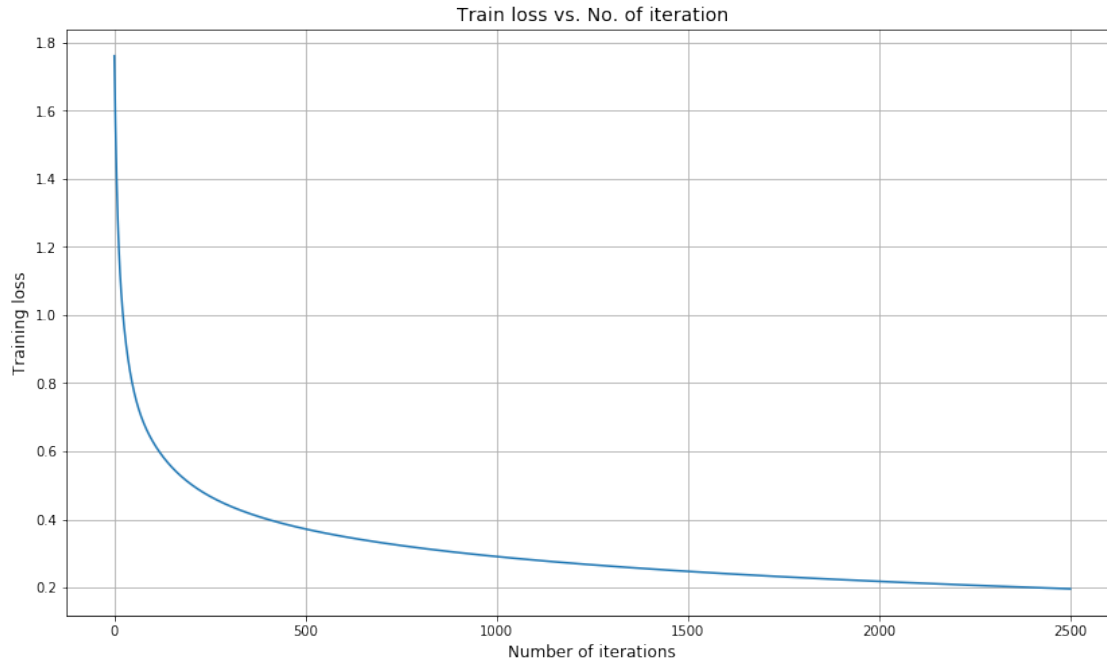
Train loss plot in the best model

```python
[37]: # plotting train_loss vs number of iteration
      plt.figure(figsize=(14, 8))

      plt.plot(train_loss)

      plt.title('Train loss vs. No. of iteration', fontsize=14)
      plt.xlabel('Number of iterations', fontsize=12)
      plt.ylabel('Training loss', fontsize=12)

      plt.grid()
      plt.show()
```

Train loss vs. No. of iteration

This shows that our `training loss` is decreses with increaing in `number of iterations`.

```
[ ]:  **Model Evaluation**
```

```
[40]:  def prediction_data(X_test, final_weights, threshold):
           predictions = sigmoid(np.matmul(X_test, final_weights))
           final_predictions = []

           for x in predictions:
               if x >= threshold:
                   final_predictions.append(1)
               else:
                   final_predictions.append(0)

           return final_predictions


       # predictions
       prediction = prediction_data(X_test, final_weights, 0.5)
```

```
[44]:  # confusion matrix function
       def get_confusion_matrix(actual, predicted):
           confusion_matrix = np.zeros((2, 2))

           for x, y in zip(actual, predicted):
               confusion_matrix[x, y] += 1
```

```
    return confusion_matrix
```

[45]:
```python
# get confusion matrix
confusion_matrix = get_confusion_matrix(y_test, prediction)

print(confusion_matrix)
```

```
[[150.  63.]
 [ 30. 318.]]
```

[46]:
```python
# accuracy
accuracy = (confusion_matrix[0][0] + confusion_matrix[1][1]) / np.
 ↪sum(confusion_matrix)

print('Accuracy: ', accuracy)
```

```
Accuracy:  0.8342245989304813
```

[47]:
```python
# recall = TP / (TP +FN)
recall = (confusion_matrix[0][0]) / (confusion_matrix[0][0] +␣
 ↪confusion_matrix[0][1])

print('Recall: ', recall)
```

```
Recall:  0.704225352112676
```

[48]:
```python
# Precision = TP / (TP + FP)
precision = (confusion_matrix[0][0]) / (confusion_matrix[0][0] +␣
 ↪confusion_matrix[1][0])

print('Precision: ', precision)
```

```
Precision:  0.8333333333333334
```

[49]:
```python
# F1-Score
f1_score = (2 * recall * precision) / (recall + precision)

print('F1 score: ', f1_score)
```

```
F1 score:  0.7633587786259541
```

[51]:
```python
# contructing ROC curve
tpr = []
fpr = []

for threshold in np.arange(0, 1, 0.1):
    prediction = prediction_data(X_test, final_weights, threshold)
```

```
    confusion_matrix = get_confusion_matrix(y_test, prediction)

    t_p_r = (confusion_matrix[0][0]) / (confusion_matrix[0][0] +␣
→confusion_matrix[0][1])
    f_p_r = (confusion_matrix[1][0]) / (confusion_matrix[1][0] +␣
→confusion_matrix[1][1])

    tpr.append(t_p_r)
    fpr.append(f_p_r)
```
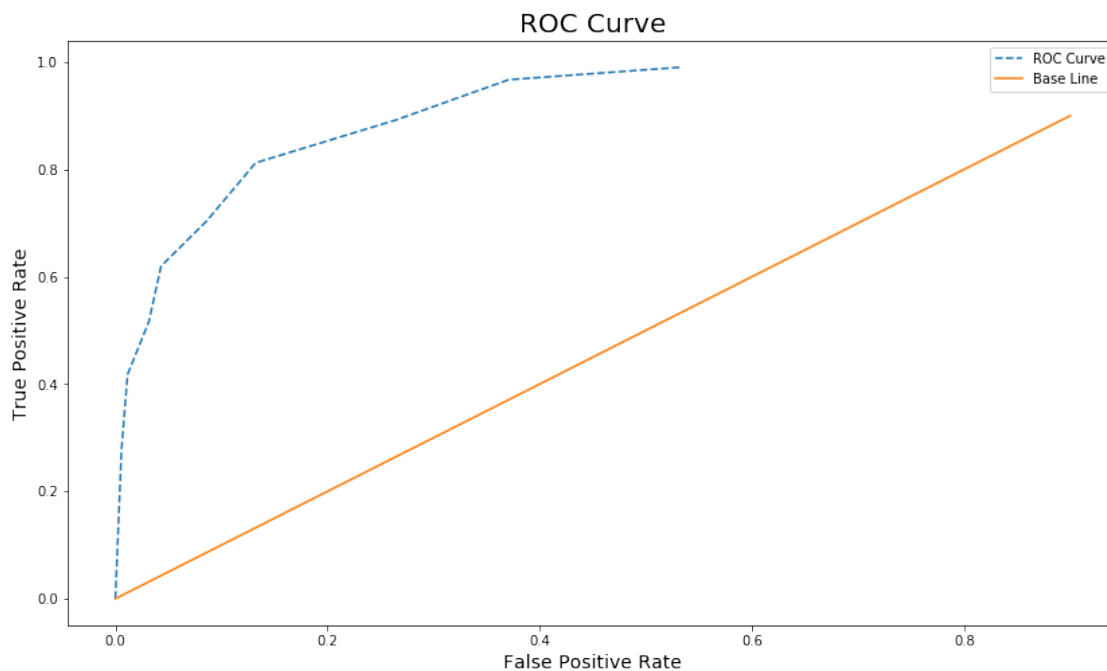
Plotting ROC curve

[52]:
```python
# ROC curve
plt.figure(figsize = (14, 8))

base = np.arange(0,1,0.1)
plt.plot(fpr, tpr, linestyle = '--', label = 'ROC Curve')
plt.plot(base, base, label = 'Base Line')

plt.title('ROC Curve', fontsize = 20)
plt.xlabel('False Positive Rate', fontsize = 14)
plt.ylabel('True Positive Rate', fontsize = 14)

plt.legend()
plt.show()
```

`[ ]:`