-------------------------------Exploratory Data Analysis- ------------------------------------- ¶

------------------------------------Import Basic Libraries----- -------------------------------------

```python
In [27]:    import pandas as pd
            import numpy as np
            import matplotlib.pyplot as plt
            import seaborn as sns
```

-------------------------------------Import both Datasets--- -------------------------------------

```python
In [30]:    lego = pd.read_csv('lego_sets.csv')
```

```python
In [31]:    lego.head(5)
```

Out[31]:

	set_num	name	year	num_parts	theme_name	parent_theme
0	00-1	Weetabix Castle	1970	471.0	Castle	Legoland
1	0011-2	Town Mini-Figures	1978	NaN	Supplemental	Town
2	0011-3	Castle 2 for 1 Bonus Offer	1987	NaN	Lion Knights	Castle
3	0012-1	Space Mini-Figures	1979	12.0	Supplemental	Space
4	0013-1	Space Mini-Figures	1979	12.0	Supplemental	Space

```python
In [32]:    theme = pd.read_csv('parent_themes.csv')
```

```python
In [38]:    theme.head(5)
```

Out[38]:

	id	name	is_licensed
0	1	Technic	False
1	22	Creator	False
2	50	Town	False
3	112	Racers	False
4	126	Space	False

```python
In [95]:    lego.columns
```

```
Out[95]:    Index(['set_num', 'name', 'year', 'num_parts', 'theme_name', 'parent_theme'],
            dtype='object')
```

```
In [96]:   ▶  theme.columns
```

Out[96]: Index(['id', 'name', 'is_licensed'], dtype='object')

```
In [97]:   ▶  lego.shape
```

Out[97]: (11986, 6)

```
In [98]:   ▶  theme.shape
```

Out[98]: (111, 3)

```
In [101]:  ▶  lego.describe().T
```

Out[101]:

	count	mean	std	min	25%	50%	75%	max
year	11986.0	2002.304772	13.471680	1950.0	1997.00	2006.0	2013.0	2017.0
num_parts	6926.0	163.782558	334.206873	-1.0	10.25	46.0	173.0	5922.0

```
In [102]:  ▶  theme.describe().T
```

Out[102]:

	count	mean	std	min	25%	50%	75%	max
id	111.0	405.27027	145.444394	1.0	298.5	411.0	505.5	610.0

```
In [104]:  ▶  lego.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11986 entries, 0 to 11985
Data columns (total 6 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   set_num       11833 non-null  object
 1   name          11833 non-null  object
 2   year          11986 non-null  int64
 3   num_parts     6926 non-null   float64
 4   theme_name    11833 non-null  object
 5   parent_theme  11986 non-null  object
dtypes: float64(1), int64(1), object(4)
memory usage: 562.0+ KB
```

```
In [105]:  ▶  theme.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 111 entries, 0 to 110
Data columns (total 3 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   id           111 non-null    int64
 1   name         111 non-null    object
 2   is_licensed  111 non-null    bool
dtypes: bool(1), int64(1), object(1)
memory usage: 2.0+ KB
```

```
In [111]:  ▶|  lego.isnull().sum()
```

Out[111]: set_num   153
           name   153
           year   0
           num_parts   5060
           theme_name   153
           parent_theme   0
           dtype: int64

```
In [113]:  ▶|  lego.duplicated().sum()
```

Out[113]: 44

```
In [114]:  ▶|  theme.duplicated().sum()
```

Out[114]: 0

```
In [112]:  ▶|  theme.isnull().sum()
```

Out[112]: id   0
           name   0
           is_licensed   0
           dtype: int64

```
In [109]:  ▶| max_df = summed_df.sort_values('num_parts', ascending=False).drop_duplicates(['
           max_df.head()
```

Out[109]:

	year	parent_theme	set_num	name_x	num_parts	
82	2017	Super Heroes	76075-1211701-130449-130521-17090470903-170902...	Wonder Woman" Warrior BattleBatmanThe MilanoTh...	13123.0	HeroesBatmanMarvelE
67	2015	Star Wars	75099-1SW911506-1SW911508-1SW911509-1SW911511-...	Rey's Speeder™SnowspeederMini Slave 1Imperial ...	11410.0	Star WarsStar Wars E
24	2007	Star Wars	7660-17659-17658-17657-17656-17661-17662-17663...	Naboo N-1 Starfighter and Vulture DroidImperia...	11361.0	Star Wars Episode 1S
59	2014	Star Wars	75046-175045-175044-175043-175039-175041-17504...	Coruscant Police GunshipRepublic AV-7 Anti-Veh...	8293.0	Star WarsStar
76	2016	Star Wars	75156-175157-175159-130279-175158-130602-13060...	Krennic's Imperial ShuttleCaptain Rex's AT-TED...	6934.0	Star Wars Roc

◀ ▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬ ▶

```
In [110]: summed_df['count'] = summed_df.groupby('year')['year'].transform('count')
          max_df = summed_df.sort_values('count', ascending=False).drop_duplicates(['year
          max_df.head()
```

Out[110]:

	year	parent_theme	set_num	name_x	num_parts	
72	2016	Disney Princess	41065-141066-141067-141068-130397-141140-14114...	Rapunzel's Best Day EverAnna & Kristoff's Slei...	1433.0	Disney PrincessDisney PrincessDis
48	2012	Super Heroes	30164-15000022-15000728-130165-130163-130160-1...	Lex LuthorHulkDC Universe Super Heroes Collect...	4843.0	BatmanAvengersConstractionAvengers
63	2015	Disney Princess	41063-141062-141061-141060-1	Ariel's Undersea PalaceElsa's Sparkling Ice Ca...	386.0	Disney PrincessDisney PrincessDis
51	2013	Minecraft	21105-121106-1	Minecraft Micro World: The VillageMinecraft Mi...	469.0	M
41	2011	SpongeBob SquarePants	3816-13815-1	Glove WorldHeroic Heroes of the Deep	168.0	SpongeBob SquarePantsSpongeE

--------------------------------Merged the dataset--

```
In [94]: merged = df.merge(theme, left_on='parent_theme', right_on='name')
```

```
In [ ]: merged.drop(columns='name_y', inplace=True)
```

```
In [43]: merged.head(5)
```

Out[43]:

	set_num	name_x	year	num_parts	theme_name	parent_theme	id	is_licensed
0	00-1	Weetabix Castle	1970	471.0	Castle	Legoland	411	False
1	00-2	Weetabix Promotional House 1	1976	NaN	Building	Legoland	411	False
2	00-3	Weetabix Promotional House 2	1976	NaN	Building	Legoland	411	False
3	00-4	Weetabix Promotional Windmill	1976	126.0	Building	Legoland	411	False
4	00-7	Weetabix Promotional Lego Village	1976	NaN	Building	Legoland	411	False

```
In [44]:   ▶|   licensed = merged[merged['is_licensed']]
               licensed.head()
```

Out[44]:

	set_num	name_x	year	num_parts	theme_name	parent_theme	id	is_licensed
3493	10018-1	Darth Maul	2001	1868.0	Star Wars	Star Wars	158	True
3494	10019-1	Rebel Blockade Runner - UCS	2001	NaN	Star Wars Episode 4/5/6	Star Wars	158	True
3495	10026-1	Naboo Starfighter - UCS	2002	NaN	Star Wars Episode 1	Star Wars	158	True
3496	10030-1	Imperial Star Destroyer - UCS	2002	3115.0	Star Wars Episode 4/5/6	Star Wars	158	True
3497	10123-1	Cloud City	2003	707.0	Star Wars Episode 4/5/6	Star Wars	158	True

```python
star_wars = licensed[licensed['parent_theme'] == 'Star Wars']
star_wars.head(50)
```

	set_num	name_x	year	num_parts	theme_name	parent_theme	id	is_licensed
3493	10018-1	Darth Maul	2001	1868.0	Star Wars	Star Wars	158	True
3494	10019-1	Rebel Blockade Runner - UCS	2001	NaN	Star Wars Episode 4/5/6	Star Wars	158	True
3495	10026-1	Naboo Starfighter - UCS	2002	NaN	Star Wars Episode 1	Star Wars	158	True
3496	10030-1	Imperial Star Destroyer - UCS	2002	3115.0	Star Wars Episode 4/5/6	Star Wars	158	True
3497	10123-1	Cloud City	2003	707.0	Star Wars Episode 4/5/6	Star Wars	158	True
3498	10129-1	Rebel Snowspeeder - UCS	2003	1456.0	Star Wars Episode 4/5/6	Star Wars	158	True
3499	10131-1	TIE Fighter Collection	2004	NaN	Star Wars Episode 4/5/6	Star Wars	158	True
3500	10134-1	Y-wing Attack Starfighter - UCS	2004	NaN	Star Wars Episode 4/5/6	Star Wars	158	True
3501	10143-1	Death Star II	2005	NaN	Star Wars Episode 4/5/6	Star Wars	158	True
3502	10144-1	Sandcrawler	2005	1679.0	Star Wars Episode 4/5/6	Star Wars	158	True
3503	10174-1	Imperial AT-ST - UCS	2006	NaN	Star Wars Episode 4/5/6	Star Wars	158	True
3504	10175-1	Vader's TIE Advanced - UCS	2006	NaN	Star Wars Episode 4/5/6	Star Wars	158	True
3505	10178-1	Motorized Walking AT-AT	2007	1137.0	Star Wars Episode 4/5/6	Star Wars	158	True
3506	10179-1	Millennium Falcon - UCS	2007	5195.0	Star Wars Episode 4/5/6	Star Wars	158	True
3507	10186-1	General Grievous	2008	1084.0	Star Wars Episode 3	Star Wars	158	True
3508	10188-1	Death Star	2008	NaN	Star Wars Episode 4/5/6	Star Wars	158	True
3509	10195-1	Republic Dropship with AT-OT	2009	NaN	Star Wars Clone Wars	Star Wars	158	True
3510	10198-1	Tantive IV	2009	NaN	Star Wars Episode 4/5/6	Star Wars	158	True
3511	10212-1	Imperial Shuttle - UCS	2010	NaN	Star Wars Episode 4/5/6	Star Wars	158	True
3512	10215-1	Obi-Wan's Jedi Starfighter - UCS	2010	NaN	Star Wars Episode 2	Star Wars	158	True
3513	10221-1	Super Star Destroyer	2011	NaN	Star Wars Episode 4/5/6	Star Wars	158	True
3514	10225-1	R2-D2	2012	2130.0	Ultimate Collector Series	Star Wars	158	True
3515	10227-1	B-wing Starfighter	2012	NaN	Star Wars Episode 4/5/6	Star Wars	158	True
3516	10236-1	Ewok Village	2013	2023.0	Star Wars Episode 4/5/6	Star Wars	158	True

	set_num	name_x	year	num_parts	theme_name	parent_theme	id	is_licensed
3517	10240-1	Red Five X-Wing Starfighter	2013	NaN	Star Wars Episode 4/5/6	Star Wars	158	True
3518	11912-1	LEGO Star Wars: Build Your Own Adventure	2016	NaN	Star Wars	Star Wars	158	True
3519	20006-1	Clone Turbo Tank - Mini	2008	64.0	Star Wars Episode 3	Star Wars	158	True
3520	20007-1	Republic Attack Cruiser - Mini	2009	NaN	Star Wars Episode 3	Star Wars	158	True
3521	20009-1	AT-TE Walker - Mini	2009	NaN	Star Wars Episode 2	Star Wars	158	True
3522	20010-1	Republic Gunship - Mini	2009	NaN	Star Wars Episode 2	Star Wars	158	True
3523	20016-1	Imperial Shuttle - Mini	2010	70.0	Star Wars Episode 4/5/6	Star Wars	158	True
3524	20018-1	AT-AT Walker - Mini	2010	NaN	Star Wars Episode 4/5/6	Star Wars	158	True
3525	20019-1	Slave I	2011	NaN	Star Wars Episode 4/5/6	Star Wars	158	True
3526	20021-1	Bounty Hunter Gunship - Mini	2011	81.0	Star Wars Clone Wars	Star Wars	158	True
3527	2853590-1	Stormtrooper	2009	NaN	Star Wars Episode 4/5/6	Star Wars	158	True
3528	2853835-1	White Boba Fett Figure	2010	NaN	Star Wars	Star Wars	158	True
3529	2856197-1	Shadow ARF Trooper	2011	NaN	Minifig Pack	Star Wars	158	True
3530	30004-1	Battle Droid on STAP	2009	NaN	Star Wars Clone Wars	Star Wars	158	True
3531	30005-1	Imperial Speeder Bike	2009	33.0	Star Wars Episode 4/5/6	Star Wars	158	True
3532	30006-1	Clone Walker	2009	31.0	Star Wars Clone Wars	Star Wars	158	True
3533	30050-1	Republic Attack Shuttle - Mini	2010	54.0	Star Wars Clone Wars	Star Wars	158	True
3534	30051-1	X-wing Fighter - Mini	2010	NaN	Star Wars Episode 4/5/6	Star Wars	158	True
3535	30052-1	AAT - Mini	2011	NaN	Star Wars Episode 2	Star Wars	158	True
3536	30053-1	Republic Attack Cruiser - Mini	2011	41.0	Star Wars Episode 3	Star Wars	158	True
3537	30054-1	AT-ST - Mini	2011	46.0	Star Wars Episode 4/5/6	Star Wars	158	True
3538	30055-1	Vulture Droid - Mini	2011	42.0	Star Wars Episode 3	Star Wars	158	True
3539	30056-1	Star Destroyer	2012	NaN	Star Wars Episode 4/5/6	Star Wars	158	True
3540	30057-1	Anakin's Pod Racer	2012	NaN	Star Wars Episode 1	Star Wars	158	True
3541	30058-1	STAP	2012	19.0	Star Wars	Star Wars	158	True
3542	30059-1	MTT	2012	51.0	Star Wars	Star Wars	158	True

```
In [53]:  ▶| merged['set_num'].isnull().sum()

Out[53]:  153

In [54]:  ▶| licensed = licensed.dropna(subset=['set_num'])

In [49]:  ▶| star_wars.shape[0]

Out[49]:  609

In [73]:  ▶| the_force = int(star_wars.shape[0]/licensed.shape[0]*100)
             the_force

Out[73]:  51

In [62]:  ▶| licensed_sorted = licensed.sort_values('year')
             licensed_sorted
```

Out[62]:

	set_num	name_x	year	num_parts	theme_name	parent_theme	id	is_licensed
3702	7161-1	Gungan Sub	1999	379.0	Star Wars Episode 1	Star Wars	158	True
3705	7171-1	Mos Espa Podrace	1999	NaN	Star Wars Episode 1	Star Wars	158	True
3690	7140-1	X-wing Fighter	1999	271.0	Star Wars Episode 4/5/6	Star Wars	158	True
3685	7130-1	Snowspeeder	1999	NaN	Star Wars Episode 4/5/6	Star Wars	158	True
3684	7128-1	Speeder Bikes	1999	93.0	Star Wars Episode 4/5/6	Star Wars	158	True
...	...	...	...	...	...	...	...	...
5692	71017-17	March Harriet™	2017	6.0	Batman	Super Heroes	482	True
5691	71017-16	Catman™	2017	NaN	Batman	Super Heroes	482	True
5690	71017-15	Zodiac Master™	2017	6.0	Batman	Super Heroes	482	True
5696	71017-20	The Mime™	2017	NaN	Batman	Super Heroes	482	True
4082	SW911725-1	Sandcrawler	2017	NaN	Star Wars	Star Wars	158	True

1179 rows × 8 columns

```
In [63]:    ▶  licensed_sorted.groupby(['year','parent_theme']).head(50)
```

Out[63]:

	set_num	name_x	year	num_parts	theme_name	parent_theme	id	is_licensed
3702	7161-1	Gungan Sub	1999	379.0	Star Wars Episode 1	Star Wars	158	True
3705	7171-1	Mos Espa Podrace	1999	NaN	Star Wars Episode 1	Star Wars	158	True
3690	7140-1	X-wing Fighter	1999	271.0	Star Wars Episode 4/5/6	Star Wars	158	True
3685	7130-1	Snowspeeder	1999	NaN	Star Wars Episode 4/5/6	Star Wars	158	True
3684	7128-1	Speeder Bikes	1999	93.0	Star Wars Episode 4/5/6	Star Wars	158	True
...	...	...	...	...	...	...	...	...
5678	70912-1	Arkham Asylum	2017	1621.0	Batman	Super Heroes	482	True
5679	70913-1	Scarecrow" Fearful Face-off	2017	NaN	Batman	Super Heroes	482	True
5680	70914-1	Bane" Toxic Truck Attack	2017	365.0	Batman	Super Heroes	482	True
5681	70915-1	Two-Face" Double Demolition	2017	563.0	Batman	Super Heroes	482	True
5682	70916-1	The Batwing	2017	1052.0	Batman	Super Heroes	482	True

1133 rows × 8 columns

```
In [80]:    ▶  licensed_sorted['count '] = 1
               licensed_sorted.head()
```

Out[80]:

	set_num	name_x	year	num_parts	theme_name	parent_theme	id	is_licensed	cou
3702	7161-1	Gungan Sub	1999	379.0	Star Wars Episode 1	Star Wars	158	True	
3705	7171-1	Mos Espa Podrace	1999	NaN	Star Wars Episode 1	Star Wars	158	True	
3690	7140-1	X-wing Fighter	1999	271.0	Star Wars Episode 4/5/6	Star Wars	158	True	
3685	7130-1	Snowspeeder	1999	NaN	Star Wars Episode 4/5/6	Star Wars	158	True	
3684	7128-1	Speeder Bikes	1999	93.0	Star Wars Episode 4/5/6	Star Wars	158	True	

◀ ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ ▶

```
In [84]:    ▶  summed_df = licensed_sorted.groupby(['year','parent_theme']).sum().reset_index(
```

Out[85]:

|	year	parent_theme	set_num	name_x	num_parts	theme_name	id	is_licens
0	1999	Star Wars	7161-17171-17140-17130-17128-17141-17121-17131...	Gungan SubMos Espa PodraceX-wing FighterSnowsp...	1384.0	Star Wars Episode 1Star Wars Episode 1Star War...	2054	
1	2000	Disney's Mickey Mouse	4165-14167-14166-14164-14178-1	Minnie's Birthday PartyMickey's MansionMickey'...	405.0	Disney's Mickey MouseDisney's Mickey MouseDisn...	1940	
2	2000	Star Wars	7134-13341-17159-17144-17124-17115-17104-14151...	A-wing FighterStar Wars #2 - Luke/Han/Boba Min...	2580.0	Star Wars Episode 4/5/6Star Wars Episode 4/5/6...	4108	
3	2001	Harry Potter	4723-14707-14722-14721-14709-14701-14702-14704...	Diagon Alley ShopsHagrid's HutGryffindorHogwar...	1284.0	Sorcerer's StoneSorcerer's StoneSorcerer's Sto...	2706	
4	2001	Star Wars	7166-17186-165028-165030-165034-18007-18008-17...	Imperial ShuttleWatto's JunkyardStar Wars Co-P...	2949.0	Star Wars Episode 4/5/6Star Wars Episode 1Star...	2212	

◀ ◻◻◻◻◻◻◻◻◻◻◻◻◻◻◻◻◻◻◻◻ ▶

1. Year-wise Distribution of LEGO Sets

```python
plt.figure(figsize=(14, 7))
sns.countplot(data=lego, x='year', color='skyblue')
plt.xticks(rotation=90)
plt.title('Year-wise Distribution of LEGO Sets')
plt.xlabel('Year')
plt.ylabel('Number of Sets Released')
plt.tight_layout()
plt.show()
```
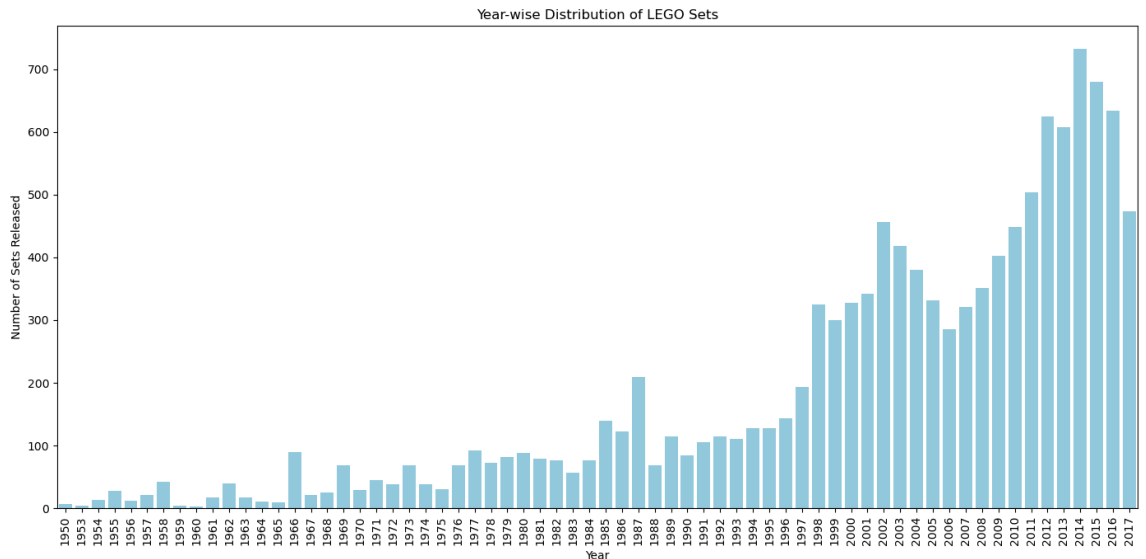

2. Top LEGO Themes by Set Count

```python
top_themes = lego['theme_name'].value_counts().head(10)
plt.figure(figsize=(14, 7))
sns.barplot(x=top_themes.index, y=top_themes.values, palette="viridis")
plt.xticks(rotation=45)
plt.title('Top LEGO Themes by Set Count')
plt.xlabel('Theme Name')
plt.ylabel('Number of Sets')
plt.tight_layout()
plt.show()
```
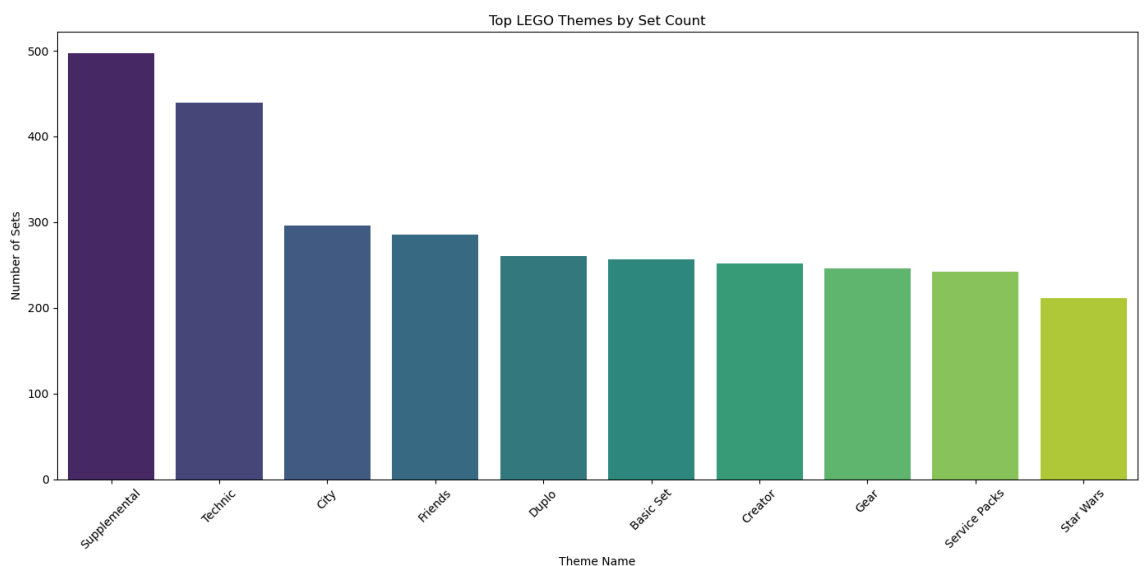
3. Top LEGO Themes by Set Count

```python
plt.figure(figsize=(14, 7))
yearly_avg_parts = lego.groupby('year')['num_parts'].mean()
sns.lineplot(x=yearly_avg_parts.index, y=yearly_avg_parts.values, color='tomato
plt.title('Year-wise Average Number of Parts in LEGO Sets')
plt.xlabel('Year')
plt.ylabel('Average Number of Parts')
plt.tight_layout()
plt.show()
```
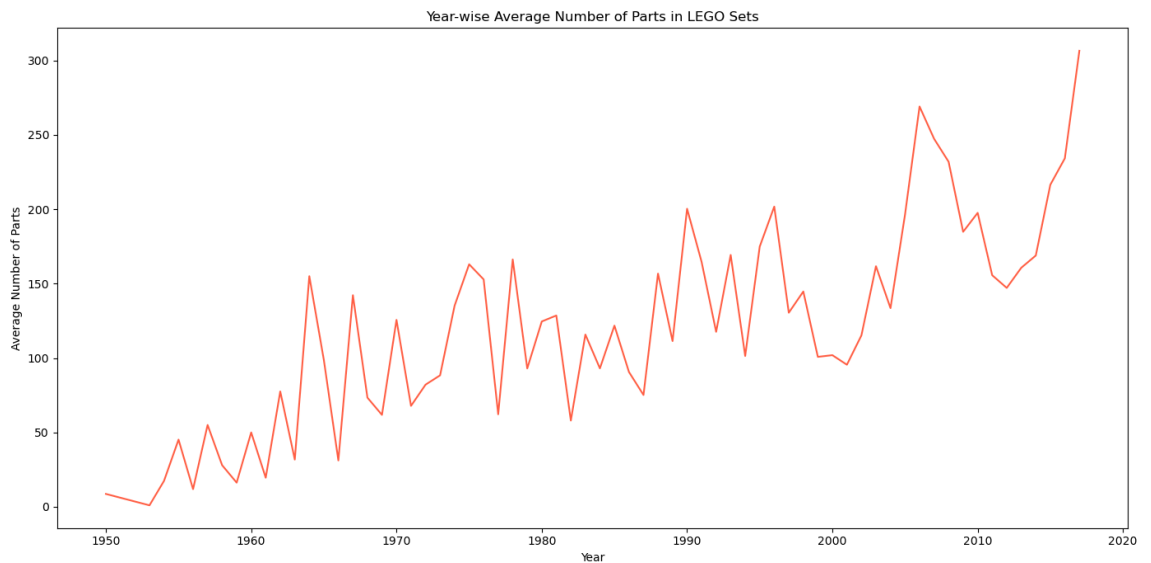

4. Distribution of Licensed vs. Non-Licensed Sets

```
plt.figure(figsize=(7, 7))
licensed_counts =merged['is_licensed'].value_counts()
plt.pie(licensed_counts, labels=['Non-Licensed', 'Licensed'], autopct='%1.1f%%'
plt.title('Distribution of Licensed vs. Non-Licensed Sets')
plt.tight_layout()
plt.show()
```
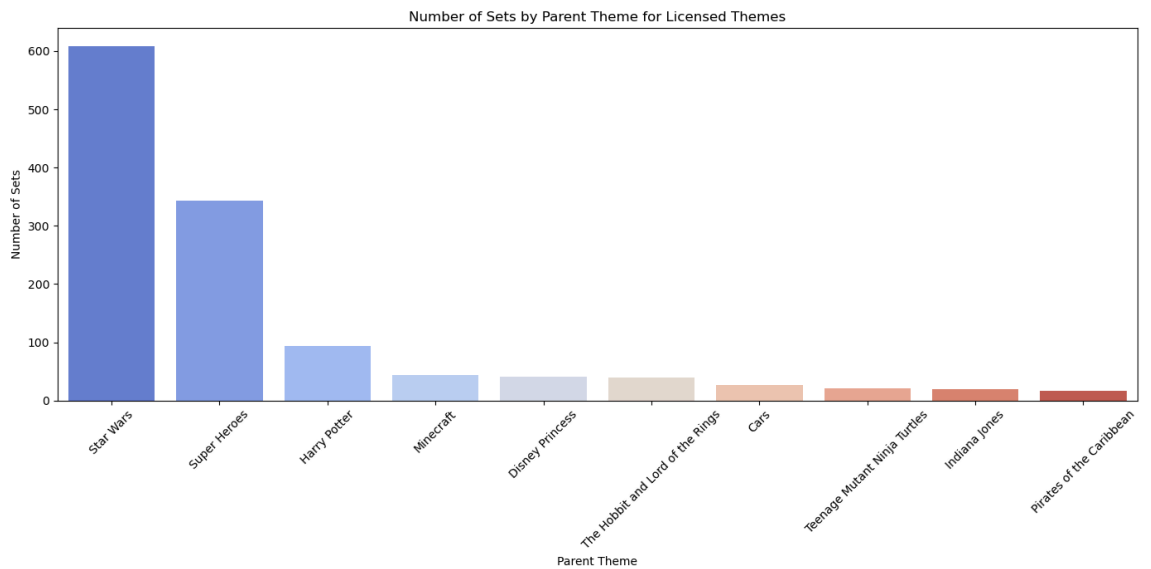
Distribution of Licensed vs. Non-Licensed Sets
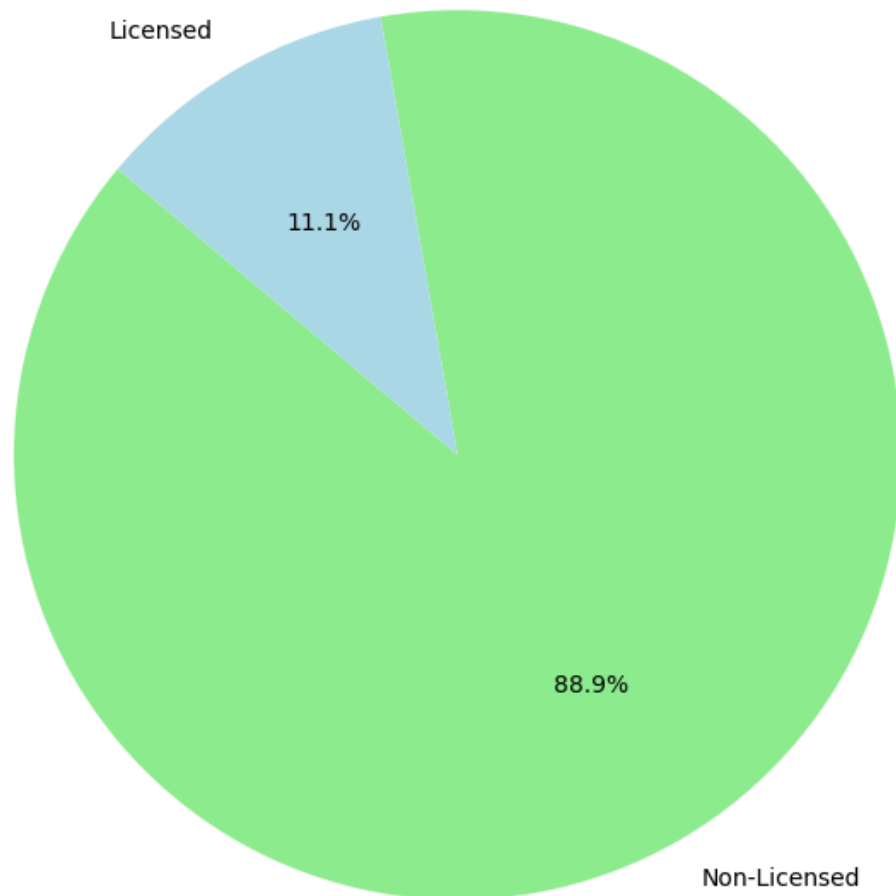
5. Number of Sets by Parent Theme for Licensed Themes Only

In [93]:

```python
licensed_sets = merged[merged['is_licensed'] == True]
sets_by_parent_theme = licensed_sets['parent_theme'].value_counts().head(10)

plt.figure(figsize=(14, 7))
sns.barplot(x=sets_by_parent_theme.index, y=sets_by_parent_theme.values, palett
plt.xticks(rotation=45)
plt.title('Number of Sets by Parent Theme for Licensed Themes')
plt.xlabel('Parent Theme')
plt.ylabel('Number of Sets')
plt.tight_layout()
plt.show()
```


Number of Sets by Parent Theme for Licensed Themes

--------------------------------------- The End --

In [ ]: