D212: Data Mining II – Task 3 Association Rules & Lift Analysis

Part I: Research Question

A1. Research Question

What items are most frequently purchased together according to the provided telecommunications dataset?

A2. Defined Goal

The goal of the market basket analysis is to identify the relationships between items that are frequently bought together. The results found can help the executive team decide whether to offer discounts on the items of interest such as bundling options and improve product placement to reduce customer churn.

Part II: Market Basket Justification

B1. Market Basket Explanation

According to Dr. Kamara's video, market basket analysis (MBA) is a data mining technique used to reveal any patterns and trends between pairs of items frequently bought together.

The expected outcome of MBA is to analyze the transaction datasets to determine which items are frequently bought together and how strong their relationships are to each other, allowing the company to utilize these findings to retain customers by offering bundling discounts and to optimize product placement on websites, in stores, or in promotions based on the item's popularity.

B2. Transaction Example

An example of a transaction from the dataset:
Apple Lightning to Digital AV Adapter, TP-Link AC1750 Smart WiFi Router, Apple Pencil

B3. Market Basket Assumption

An assumption of market basket analysis is there is some type of relationship between a pair of items in the large dataset. This once again refers to the "if-then" statement as mentioned above in part B2.

Part III: Data Preparation & Analysis

C1. Transforming the Dataset

A copy of the cleaned data set has been attached as "df_clean1.csv".

```python
# Import the necessary packages & Libraries
import pandas as pd
from pandas import DataFrame
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
from mlxtend.frequent_patterns import apriori
from mlxtend.frequent_patterns import association_rules
from mlxtend.preprocessing import TransactionEncoder

# Ignore warning code
import warnings
warnings.filterwarnings('ignore')

# Load the data set into the pandas data frame by using read_csv command
data = pd.read_csv(r'C:\Users\ashle\Downloads\D212\teleco_market_basket.csv')
data.head(5)
```

Out[1]:

	Item01	Item02	Item03	Item04	Item05	Item06	Item07	Item08	Item09	Item10	Item11	Item12	Item13	Item14	Item15	Item16	Item17	Item18	Item19	Item20
0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	Logitech M510 Wireless mouse	HP 63 Ink	HP 65 ink	nonda USB C to USB Adapter	10ft iPhone Charger Cable	HP 902XL ink	Creative Pebble 2.0 Speakers	Cleaning Gel Universal Dust Cleaner	Micro Center 32GB Memory card	YUNSONG 3pack 6ft Nylon Lightning Cable	TopMate C5 Laptop Cooler pad	Apple USB-C Charger cable	HyperX Cloud Stinger Headset	TONOR USB Gaming Microphone	Dust-Off Compressed Gas 2 pack	3A USB Type C Cable 3 pack 6FT	HOVAMP iPhone charger	SanDisk Ultra 128GB card	FEEL2NICE 5 pack 10ft Lighning cable	FEIYOLD Blue light Blocking Glasses
2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	Apple Lightning to Digital AV Adapter	TP-Link AC1750 Smart WiFi Router	Apple Pencil	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

```python
data.shape
```

Out[2]: (15002, 20)

```python
# Remove blank rows from the dataset
data = data[data['Item01'].notna()]
data.shape
```

Out[3]: (7501, 20)

```python
# Dataframe converted to list of Lists
rows = []
for i in range (0, 7501):
    rows.append([str(data.values[i,j]) for j in range(0,20)])
```

```python
# Create an array
DE = TransactionEncoder()
array = DE.fit(rows).transform(rows)

# Convert the array to dataframe
transaction = pd.DataFrame(array, columns = DE.columns_)
```

```python
# Display the transaction (True means there is a missing value)
transaction
```

Out[6]:

	10ft iPhone Charger Cable	10ft iPhone Charger Cable 2 Pack	3 pack Nylon Braided Lightning Cable	3A USB Type C Cable 3 pack 6FT	5pack Nylon Braided USB C cables	ARRIS SURFboard SB8200 Cable Modem	Anker 2-in-1 USB Card Reader	Anker 4-port USB hub	Anker USB C to HDMI Adapter	Apple Lightning to Digital AV Adapter	...	iFixit Pro Tech Toolkit	iPhone 11 case	iPhone 12 Charger cable	iPhone 12 Pro case	iPhone 12 case	iPhone Charger Cable Anker 6ft	iPhone SE case	nan	nonda USB C to USB Adapter	seenda Wireless mouse
0	True	False	False	True	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	True	False
1	False	False	False	False	False	False	False	False	False	True	...	False	False	False	False	False	False	False	True	False	False
2	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	True	False	False
3	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	True	False	False
4	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	True	False	False
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
7496	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	True	False	False
7497	False	False	False	False	False	True	False	False	False	True	...	False	False	False	False	False	False	False	True	False	False
7498	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	True	False	False
7499	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	True	False	False
7500	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	True	False	False

7501 rows × 120 columns

```python
# Display items as columns
for col in transaction.columns:
    print(col)
```

```
10ft iPhone Charger Cable
10ft iPhone Charger Cable 2 Pack
3 pack Nylon Braided Lightning Cable
3A USB Type C Cable 3 pack 6FT
5pack Nylon Braided USB C cables
ARRIS SURFboard SB8200 Cable Modem
Anker 2-in-1 USB Card Reader
Anker 4-port USB hub
Anker USB C to HDMI Adapter
Apple Lightning to Digital AV Adapter
Apple Lightning to USB cable
Apple Magic Mouse 2
Apple Pencil
Apple Pencil 2nd Gen
Apple Power Adapter Extension Cable
Apple USB-C Charger cable
AutoFocus 1080p Webcam
BENGOO G9000 headset
Blue Light Blocking Glasses
Blue Light Blocking Glasses 2pack
Brother Genuine High Yield Toner Cartridge
Cat 6 Ethernet Cable 50ft
Cat8 Ethernet Cable
CicTsing MM057 2.4G Wireless Mouse
Cleaning Gel Universal Dust Cleaner
Creative Pebble 2.0 Speakers
DisplayPort ot HDMI adapter
Dust-OFF Compressed Gas
Dust-Off Compressed Gas 2 pack
FEEL2NICE 5 pack 10ft Lighning cable
FEIYOLD Blue light Blocking Glasses
Falcon Dust Off Compressed Gas
HOVAMP MFi 6pack Lightning Cable
HOVAMP iPhone charger
HP 61 2 pack ink
HP 61 Tri-color ink
HP 61 ink
HP 62XL Tri-Color ink
```

```
HP 62XL ink
HP 63 Ink
HP 63 Tri-color ink
HP 63XL Ink
HP 63XL Tri-color ink
HP 64 Tri-Color ink
HP 64 ink
HP 65 ink
HP 902XL ink
HP 952 ink
HP ENVY 5055 printer
HP952XL ink
HooToo USB C Hub
HyperX Cloud Stinger Headset
Jelly Comb 2.4G Slim Wireless mouse
Leader Desk Pad Protector
Logitech M510 Wireless mouse
Logitech MK270 Wireless Keyboard/Mouse
Logitech MK345 Wireless combo
Logitech USB H390 headset
M.2 Screw kit
Mfi-Certified Lightning to USB A Cable
Micro Center 32GB Memory card
Microsot Surface Dock 2
Moread HDMI to VGA Adapter
Mpow HC6 USB Headset
NETGEAR CM500 Cable Modem
NETGEAR Nighthawk WiFi Router
NETGEAR Orbi Home Mesh WiFi System
Nylon Braided Lightning to USB cable
PS4 Headset
Premium Nylon USB Cable
RUNMUS Gaming Headset
SAMSUNG 128GB card
SAMSUNG 256 GB card
SAMSUNG EVO 32GB card
SAMSUNG EVO 64GB card
Sabrent 4-port USB 3.0 hub
SanDisk 128GB Ultra microSDXC card
SanDisk 128GB card
SanDisk 128GB microSDXC card
SanDisk 32GB Ultra SDHC card
SanDisk 32GB card
SanDisk Extreme 128GB card
SanDisk Extreme 256GB card
SanDisk Extreme 32GB 2pack card
SanDisk Extreme Pro 128GB card
SanDisk Extreme Pro 64GB card
SanDisk Ultra 128GB card
SanDisk Ultra 256GB card
SanDisk Ultra 400GB card
SanDisk Ultra 64GB card
Screen Mom Screen Cleaner kit
Stylus Pen for iPad
Syntech USB C to USB Adapter
TONOR USB Gaming Microphone
TP-Link AC1750 Smart WiFi Router
TP-Link AC4000 WiFi router
TopMate C5 Laptop Cooler pad
UNEN Mfi Certified 5-pack Lightning Cable
USB 2.0 Printer cable
USB C to USB Male Adapter
USB Type C Cable
USB Type C to USB-A Charger cable
VIVO Dual LCD Monitor Desk mount
VicTsing Mouse Pad
VicTsing Wireless mouse
Vsco 70 pack stickers
Webcam with Microphone


XPOWER A-2 Air Pump blower
YUNSONG 3pack 6ft Nylon Lightning Cable
hP 65 Tri-color ink
iFixit Pro Tech Toolkit
iPhone 11 case
iPhone 12 Charger cable
iPhone 12 Pro case
iPhone 12 case
iPhone Charger Cable Anker 6ft
iPhone SE case
nan
nonda USB C to USB Adapter
seenda Wireless mouse
```

In [8]:
```python
# Remove NaN columns from the transformed dataset
cleaned_df = transaction.drop(['nan'], axis = 1)
cleaned_df.head(7501)
```

Out[8]:

	10ft iPhone Charger Cable	10ft iPhone Charger Cable 2 Pack	3 pack Nylon Braided Lightning Cable	3A USB Type C Cable 3 pack 6FT	5pack Nylon Braided USB C cables	ARRIS SURFboard SB8200 Cable Modem	Anker 2-in-1 USB Card Reader	Anker 4-port USB hub	Anker USB C to HDMI Adapter	Apple Lightning to Digital AV Adapter	...	hP 65 Tri-color ink	iFixit Pro Tech Toolkit	iPhone 11 case	iPhone 12 Charger cable	iPhone 12 Pro case	iPhone 12 case	iPhone Charger Cable Anker 6ft	iPhone SE case	nonda USB C to USB Adapter	seenda Wireless mouse
0	True	False	False	True	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	True	False
1	False	False	False	False	False	False	False	False	False	True	...	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
7496	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False
7497	False	False	False	False	False	True	False	False	False	True	...	False	False	False	False	False	False	False	False	False	False
7498	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False
7499	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False
7500	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False

7501 rows × 119 columns

In [9]:
```python
# Write the transformed dataset to csv
cleaned_df.to_csv('df_clean1.csv', index=False)
cleaned_df.columns
```

Out[9]:
```
Index(['10ft iPhone Charger Cable', '10ft iPhone Charger Cable 2 Pack',
       '3 pack Nylon Braided Lightning Cable',
       '3A USB Type C Cable 3 pack 6FT', '5pack Nylon Braided USB C cables',
       'ARRIS SURFboard SB8200 Cable Modem', 'Anker 2-in-1 USB Card Reader',
       'Anker 4-port USB hub', 'Anker USB C to HDMI Adapter',
       'Apple Lightning to Digital AV Adapter',
       ...
       'hP 65 Tri-color ink', 'iFixit Pro Tech Toolkit', 'iPhone 11 case',
       'iPhone 12 Charger cable', 'iPhone 12 Pro case', 'iPhone 12 case',
       'iPhone Charger Cable Anker 6ft', 'iPhone SE case',
       'nonda USB C to USB Adapter', 'seenda Wireless mouse'],
      dtype='object', length=119)
```

C2. Code Execution

```
In [10]: # Loading the transformed dataset
         data = pd.read_csv('df_clean1.csv')
         data.head()
```
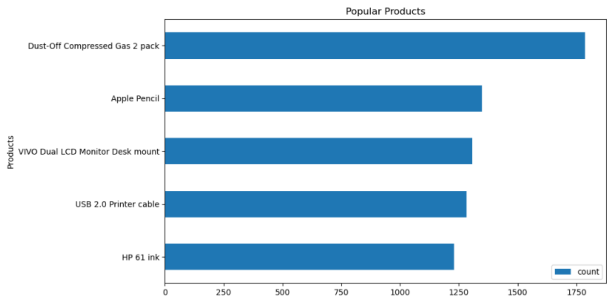
	10ft iPHone Charger Cable	10ft iPHone Charger Cable 2 Pack	3 pack Nylon Braided Lightning Cable	3A USB Type C Cable 3 pack 6FT	5pack Nylon Braided USB C cables	ARRIS SURFboard SB8200 Cable Modem	Anker 2-in-1 USB Card Reader	Anker 4-port USB hub	Anker USB C to HDMI Adapter	Apple Lightning to Digital AV Adapter	...	hP 65 Tri-color ink	iFixit Pro Tech Toolkit	iPhone 11 case	iPhone 12 Charger cable	iPhone 12 Pro case	iPhone 12 case	iPhone Charger Cable Anker 6ft	iPhone SE case	nonda USB C to USB Adapter	seenda Wireless mouse
0	True	False	False	True	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	True	False
1	False	False	False	False	False	False	False	False	False	True	...	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False

5 rows × 119 columns

```
In [11]: # The five most popular products in the transformed dataset
         count = data.loc[:,:].sum()
         pop_item = count.sort_values(ascending = False).head(5)
         pop_item = pop_item.to_frame()
         pop_item = pop_item.reset_index()
         pop_item = pop_item.rename(columns = {'index': 'Products',0:'count'})

         # Data visualization of the most popular items
         plt.rcParams['figure.figsize'] = (10,6)
         ax = pop_item.plot.barh(x = 'Products', y = 'count')
         plt.title('Popular Products')
         plt.gca().invert_yaxis()
```



```
In [12]: # Create the Apriori object called rules
         rules = apriori(data, min_support = 0.02, use_colnames = True)
         rules.head(5)
```

Out[12]:	support	itemsets
0	0.050527	(10ft iPHone Charger Cable 2 Pack)
1	0.042528	(3A USB Type C Cable 3 pack 6FT)
2	0.029463	(Anker 2-in-1 USB Card Reader)
3	0.068391	(Anker USB C to HDMI Adapter)
4	0.087188	(Apple Lightning to Digital AV Adapter)

C3. Association Rules Table

```
In [13]: # Create the rules table (lift > 1, filtering on top three values)
         rules_table = association_rules(rules, metric='lift', min_threshold=1)
         rules_table.head(10)
```

Out[13]:	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
0	{10ft iPHone Charger Cable 2 Pack)	(Dust-Off Compressed Gas 2 pack)	0.050527	0.238368	0.023064	0.456464	1.914955	0.011020	1.401255	0.503221
1	(Dust-Off Compressed Gas 2 pack)	(10ft iPHone Charger Cable 2 Pack)	0.238368	0.050527	0.023064	0.096756	1.914955	0.011020	1.051182	0.627330
2	(Anker USB C to HDMI Adapter)	(Dust-Off Compressed Gas 2 pack)	0.068391	0.238368	0.024397	0.356725	1.496530	0.008095	1.183991	0.356144
3	(Dust-Off Compressed Gas 2 pack)	(Anker USB C to HDMI Adapter)	0.238368	0.068391	0.024397	0.102349	1.496530	0.008095	1.037830	0.435627
4	(VIVO Dual LCD Monitor Desk mount)	(Anker USB C to HDMI Adapter)	0.174110	0.068391	0.020931	0.120214	1.757755	0.009023	1.058905	0.521973
5	(Anker USB C to HDMI Adapter)	(VIVO Dual LCD Monitor Desk mount)	0.068391	0.174110	0.020931	0.306075	1.757755	0.009023	1.190117	0.462740
6	(Apple Lightning to Digital AV Adapter)	(Apple Pencil)	0.087188	0.179709	0.028796	0.330275	1.837830	0.013128	1.224818	0.499424
7	(Apple Pencil)	(Apple Lightning to Digital AV Adapter)	0.179709	0.087188	0.028796	0.160237	1.837830	0.013128	1.086988	0.555754
8	(Apple Lightning to Digital AV Adapter)	(Dust-Off Compressed Gas 2 pack)	0.087188	0.238368	0.024397	0.279817	1.173883	0.003614	1.057552	0.162275
9	(Dust-Off Compressed Gas 2 pack)	(Apple Lightning to Digital AV Adapter)	0.238368	0.087188	0.024397	0.102349	1.173883	0.003614	1.016889	0.194486

C4. Top Three Rules

A screenshot of the top three relevant rules can be found below.

```
In [14]: # Sort the rules by the confidence metric
         top_three_rules = rules_table.sort_values('confidence', ascending=False).head(3)
         top_three_rules
```

Out[14]:	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
0	(10ft iPHone Charger Cable 2 Pack)	(Dust-Off Compressed Gas 2 pack)	0.050527	0.238368	0.023064	0.456464	1.914955	0.011020	1.401255	0.503221
36	(FEIYOLD Blue light Blocking Glasses)	(Dust-Off Compressed Gas 2 pack)	0.065858	0.238368	0.027596	0.419028	1.757904	0.011898	1.310962	0.461536
52	(SanDisk Ultra 64GB card)	(Dust-Off Compressed Gas 2 pack)	0.098254	0.238368	0.040928	0.416554	1.747522	0.017507	1.305401	0.474369

```
In [15]: # Sort the rules by the lift metric
         top_three_rules = rules_table.sort_values('lift', ascending=False).head(3)
         top_three_rules
```

Out[15]:	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
85	(VIVO Dual LCD Monitor Desk mount)	(SanDisk Ultra 64GB card)	0.174110	0.098254	0.039195	0.225115	2.291162	0.022088	1.163716	0.682343
84	(SanDisk Ultra 64GB card)	(VIVO Dual LCD Monitor Desk mount)	0.098254	0.174110	0.039195	0.398915	2.291162	0.022088	1.373997	0.624943
64	(FEIYOLD Blue light Blocking Glasses)	(VIVO Dual LCD Monitor Desk mount)	0.065858	0.174110	0.022930	0.348178	1.999758	0.011464	1.267048	0.535186

```
In [16]: # Sort the rules by the support metric
         top_three_rules = rules_table.sort_values('support', ascending=False).head(3)
         top_three_rules
```

Out[16]:	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
62	(VIVO Dual LCD Monitor Desk mount)	(Dust-Off Compressed Gas 2 pack)	0.174110	0.238368	0.059725	0.343032	1.439085	0.018223	1.159314	0.369437
63	(Dust-Off Compressed Gas 2 pack)	(VIVO Dual LCD Monitor Desk mount)	0.238368	0.174110	0.059725	0.250559	1.439085	0.018223	1.102008	0.400606
41	(Dust-Off Compressed Gas 2 pack)	(HP 61 ink)	0.238368	0.163845	0.052660	0.220917	1.348332	0.013604	1.073256	0.339197

Part IV: Data Summary & Implications

D1. Significance of Support, Life, & Confidence Summary

In this analysis, the significance of support, life, and confidence is that these are the key metrics used to eliminate uncommon associations and focus on the more common, popular associations between items in the dataset. The descriptions of the metrics can be found below:

- Support: The support metric measures how often the item sets appear in the dataset together. A higher number represents a popular item set. According to the analysis, VIVO Dual LCD Monitor Desk Mount and the Dust-Off Compressed Gas 2 pack appear the most frequent together.
- Confidence: The confidence metric measures the strength of an item being purchased with another item, meaning how likely the item combination is to be purchased together. The higher the confidence, the stronger the association between the items. The pair with the highest confidence metric according to the analysis would be the 10ft iPhone Charger Cable 2 Pack and the Dust-Off Compressed Gas 2 pack.
- Lift: The lift metric is a ratio of the actual support to the expected support of the pair of items if they were independent. Essentially, this shows if the pair of items are purchased more together than they are purchased separately. A value greater than 1 represents the items are associated with each other and the first item has some type of influence on the purchase of the second item.

D2. Practical Significance of Findings

The practical significance of the analysis shows that the relationships between item sets can be found using the market basket analysis. Based on the association rules discussed in D1, the VIVO Dual LCD Monitor Desk Mount and the Dust-Off Compressed Gas 2 pack have a high support value, indicating they are a popular item set, commonly purchased together. The VIVO Dual LCD Monitor Desk mount and SanDisk Ultra 64GB card have a 2.29 lift value, much greater than 1. This indicates the first item has some type of influence on the purchase of the second item. Therefore, I believe the company could benefit from offering bundling discounts on items frequently purchased together and enhancing product placement to retain customers and increase customer satisfaction.

D3. Course of Action

The VIVO Dual LCD Monitor Desk Mount and the Dust-Off Compressed Gas 2 pack appear the most frequent together and are in the top three most popular products. The SanDisk Ultra 64GB card and FEIYOLD Blue Light Blocking Glasses are not as popular but are paired up frequently with the monitor and desk mount. To increase sales for the less popular items, I recommend adding some bundling discounts with the more popular items. For example, if you purchase the mount (full price), you can get the blue light glasses for 30% off. I also recommend arranging the items on websites or in stores to where they are near each other and list it under "frequently bought together" to increase more foot/site traffic.

Part V: Attachments

E. <u>Panopto Video of Code/Programs</u>

F. <u>Sources for Third-Party Code</u>

Kamara, K. (n.d.). Market Basket Analysis in Python. WGU Panopto. Retrieved January 2, 2025, from
https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=db85c4f1-0da5-4bde-a1a4-b07c0019d46d.

G. <u>Sources</u>

Kamara, K. (n.d.). Market Basket Analysis in Theory. WGU Panopto. Retrieved January 2, 2025, from
https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=9541a29b-2f14-4c5d-9d86-af030005bcf6.

H. <u>Professional Communication</u>