# *Instructor Prep-sheet:*

# **Leveraging Generative AI for Data Analysts Instructor Demo**

## **TL;DR:**

This prep-sheet is designed to give you an idea of how to conduct your live ChatGPT demonstration. A few things to remember:

1. Do a dry run of your demo before you are live and in front of students
2. Be prepared for slight variance in ChatGPT’s outputs each time you do the demo
3. Highlight the impressive capabilities of ChatGPT, along with its limitations!

* *If you end up with a demo that was particularly interesting, please share the link* [*here*](https://docs.google.com/spreadsheets/d/1eWa3s_gkPYEfLKoBXG-83LErmMJ60elgCjuePFlPo4Y/edit#gid=0) *along with a short description!*

## **Sample Materials:**

* [Sample ChatGPT Chat Log](https://chat.openai.com/share/628c8f41-33e6-42fd-ab1f-620ec0f713fc)
* **Data**: First 30 rows of Orders.csv from Superstore (**excluding** postal\_code and region\_id)
  + [*Here*](https://drive.google.com/file/d/18pIHqpn3PZrPgfuQACkMk5pKb5MpSuq9/view?usp=drive_link) *is a separate .csv with just the data you will need*

## **Preparation:**

* Review the [sample ChatGPT thread](https://chat.openai.com/share/628c8f41-33e6-42fd-ab1f-620ec0f713fc) in **detail**
  + This is an example of how you can structure the demo.
  + It is recommended to start the demo with the same prompt every time.
  + At the end of the day this is your demo - take it in whatever direction you find most impactful based on your students and teaching style
* Login to [ChatGPT](https://chat.openai.com/) (you will just use the free version for this demo)
* Download and open the [sample data](https://drive.google.com/file/d/18pIHqpn3PZrPgfuQACkMk5pKb5MpSuq9/view?usp=drive_link)

## **Instructor Demo Outline:**

[**Sample ChatGPT Thread**](https://chat.openai.com/share/628c8f41-33e6-42fd-ab1f-620ec0f713fc)

1. **Introduce The Data:**
   1. First 30 rows of Orders.csv from Superstore (excluding postal\_code and region\_id)
   2. We are only looking at 30 rows of data, but the free version of ChatGPT can’t handle much more
2. **Discuss Some Limitations:**
   1. 30 rows of sales data is basically nothing…
   2. ***You should never put any confidential or sensitive information into ChatGPT*** - discuss the implications of this when it comes to things like consumer data, proprietary info, etc.
3. **Enter The First Prompt:**
   1. ***Act as a data analyst. Can you Create a data dictionary for this data?  
        
      \*copy and paste the sample data\****
   2. If necessary, write a second prompt and ask ChatGPT to give you the data dictionary in table form.

*****From here, you can follow the* [*sample thread*](https://chat.openai.com/share/628c8f41-33e6-42fd-ab1f-620ec0f713fc) *closely, or take the demo in a direction you feel will benefit your class the most. Either way, try not to spend more than 15-20 minutes total on the demonstration.*

****

1. **Demonstrate “laddering” by asking ChatGPT more specific questions about the recommendations it made for analysis**
   1. Pause to discuss ChatGPT’s outputs
   2. Demonstrate how a good data analyst should stay critical while using the tool
2. **Ask ChatGPT to write SQL Queries:**
   1. Analyze the output carefully
   2. Be critical and conscious of things ChatGPT might “make up”
      1. *At the end of the* [*thread*](https://chat.openai.com/share/628c8f41-33e6-42fd-ab1f-620ec0f713fc)*, ChatGPT selects “product\_category” even though that is not a field in the data you provided.*
3. **Wrap up Discussion - What have we learned?**
   1. Data Dictionaries: ChatGPT can save data analysts a lot of time creating data dictionaries
   2. Ideas for where to take your analysis
   3. Starting point for SQL queries
   4. ChatGPT will do exactly what you ask it to do, even if it has to make things up and make assumptions

## **Part Two** (Later on during a Python/Pandas lesson) **:**

Pickup the superstore thread where you left off, and start using Pandas ([sample ChatGPT thread](https://chat.openai.com/share/80db555d-fafd-4ec1-89f4-d14d61d04396)).

Here is a high level breakdown of the sample thread linked above:

* In the sample thread, the user asks ChatGPT to put that data into a Pandas dataframe.
* From there, the user asks ChatGPT to create a few visualizations
* The first couple of examples seem to be working fine, but eventually, the user gets an error
* A skilled data analyst would notice something is “off” about the error - the code references “Category” when filtering the dataframe
* However, there was no “Category” column in the original data! Somewhere along the line, ChatGPT generated the “Category” column on its own!

**Note**: Working backwards from the error to figure out what went wrong would be an incredibly powerful activity to do in class. It would also drive home the point that ChatGPT is only as good as its user! A novice data analyst would likely struggle to find the root cause of the error, and may fail to notice that the integrity of their data had been compromised by ChatGPT!

## **Other Use Cases to Share:**

* Maps
* SQL Queries
  + Address Parsing/Address Standardization
* Excel Formulas
* Commenting Code/Docstrings
* Optimizing Code (SQL, Python, Pandas, etc.)

## **Orders Data:**

order\_id order\_date ship\_date ship\_mode customer\_id product\_id sales quantity discount profit

CA-2016-2608109 2016-09-17 00:00:00 2016-09-17 00:00:00 Same Day TT-21070 OFF-FA-10002280 45.00 9 0.00 0.00

ES-2018-1945518 2018-08-29 00:00:00 2018-09-02 00:00:00 Same Day AC-10420 OFF-FA-10001724 114.00 10 0.03 0.00

ES-2018-853948 2018-07-28 00:00:00 2018-07-29 00:00:00 Same Day AC-10420 OFF-FA-10001724 114.00 10 0.03 0.00

ES-2019-1912507 2019-04-01 00:00:00 2019-04-05 00:00:00 Same Day MD-17860 OFF-PA-10001758 114.00 4 0.05 0.00

ES-2019-2038080 2019-03-25 00:00:00 2019-03-28 00:00:00 Same Day MD-17860 OFF-PA-10001758 114.00 4 0.05 0.00

ID-2019-4299999 2019-05-28 00:00:00 2019-06-01 00:00:00 Same Day EM-14095 OFF-PA-10003654 76.00 5 0.05 0.00

MX-2018-2938153 2018-10-24 00:00:00 2018-10-27 00:00:00 Same Day MC-17635 OFF-FA-10001104 54.00 6 0.03 0.00

MX-2018-3389872 2018-12-09 00:00:00 2018-12-11 00:00:00 Same Day KN-16450 OFF-LA-10000633 33.00 6 0.03 0.00

MX-2018-3434064 2018-09-30 00:00:00 2018-10-02 00:00:00 Same Day AG-10390 OFF-FA-10000540 44.00 8 0.03 0.00

MX-2018-4087259 2018-07-11 00:00:00 2018-07-12 00:00:00 Same Day ML-18265 OFF-FA-10000467 46.00 5 0.03 0.00

MX-2019-3973159 2019-09-17 00:00:00 2019-09-17 00:00:00 Same Day JE-15745 OFF-LA-10000633 11.00 2 0.05 0.00

MX-2019-5358330 2019-07-19 00:00:00 2019-07-21 00:00:00 Same Day MO-17500 OFF-PA-10000591 166.00 5 0.05 0.00

AJ-2016-1487715 2016-12-08 00:00:00 2016-12-09 00:00:00 Second Class JH-15820 OFF-ROG-10002682 24.45 1 0.30 0.30

AO-2019-4862102 2019-08-21 00:00:00 2019-08-23 00:00:00 Standard Class AZ-10750 FUR-RUB-10001796 34.92 2 0.10 0.05

BO-2019-1355783 2019-04-16 00:00:00 2019-04-19 00:00:00 Second Class MS-17980 OFF-ELD-10002578 20.58 2 0.35 0.28

CA-2016-1100747 2016-11-05 00:00:00 2016-11-08 00:00:00 Same Day NG-18355 OFF-PA-10004082 15.96 2 0.20 0.20

CA-2015-1688572 2015-12-26 00:00:00 2015-12-27 00:00:00 Second Class VS-21820 OFF-BI-10003708 10.43 7 0.30 0.30

CA-2016-100006 2016-09-07 00:00:00 2016-09-13 00:00:00 Standard Class DK-13375 TEC-PH-10002075 377.97 3 0.20 109.61

CA-2016-100293 2016-03-14 00:00:00 2016-03-18 00:00:00 Standard Class NF-18475 OFF-PA-10000176 91.06 6 0.00 31.87

CA-2016-1003030 2016-09-30 00:00:00 2016-10-04 00:00:00 Same Day BN-11470 OFF-PA-10001800 10.37 2 0.00 0.00

CA-2016-100678 2016-04-18 00:00:00 2016-04-22 00:00:00 Standard Class KM-16720 FUR-CH-10002602 317.06 3 0.35 -18.12

CA-2016-100762 2016-11-24 00:00:00 2016-11-29 00:00:00 Standard Class NG-18355 OFF-AR-10000380 151.92 4 0.30 45.58

CA-2016-100895 2016-06-02 00:00:00 2016-06-06 00:00:00 Standard Class SV-20785 OFF-ST-10001490 356.94 2 0.05 107.08

IT-2019-5656576 2019-11-27 00:00:00 2019-12-01 00:00:00 Same Day AB-10105 OFF-LA-10000379 11.34 2 0.05 0.00

CA-2016-100916 2016-10-21 00:00:00 2016-10-26 00:00:00 Standard Class FH-14275 OFF-ST-10003479 194.70 5 0.05 9.73

CA-2016-1012398 2016-08-25 00:00:00 2016-08-27 00:00:00 Standard Class BM-11650 OFF-AP-10000358 25.96 2 0.50 0.50

CA-2016-1012695 2016-12-09 00:00:00 2016-12-12 00:00:00 Second Class KD-16270 OFF-BI-10001267 1.23 1 0.30 0.30

CA-2016-1013462 2016-12-28 00:00:00 2016-12-28 00:00:00 Same Day AJ-10795 OFF-LA-10001474 274.77 9 0.20 0.20

CA-2016-101364 2016-12-22 00:00:00 2016-12-26 00:00:00 Standard Class TW-21025 OFF-BI-10003984 296.71 13 0.20 100.14