\Slides:
https://www.canva.com/design/DAFyI_z5e4M/kaqKr9h5XX1vVXkIiiOu3g/edit?utm_content=DAFyI_z5e4M&utm_campaign=designshare&utm_medium=link2&utm_source=sharebutton

History and Evolution
- how data and analytics is used within the company.
  - Content Recommendations
  - Marketing
  - https://getstarted.tiktok.com/ttvalue?attr_source=google&attr_medium=search-br-ad&attr_adgroup_id=131664211377&attr_term=tiktok%20advertising&gad=1&gclid=Cj0KCQjw4vKpBhCZARIsAOKHoWTAvUXwMyAURXcKwgVTdSA0rl1qlMGmG8dJDGiI9djX8dgjI9yBy7EaAIpeEALw_wcB&lang=en
- Did the company start with or evolve into a data driven approach to developing products and services?
  - TikTok has been utilizing a data-driven approach to develop its products and services right from the start. The platform's primary purpose is to recommend videos to its users, and to achieve this, TikTok relies on a sophisticated algorithm that is driven by data. This algorithm takes into account various data points to determine the most relevant videos for each user.
  - For example, the algorithm might consider factors such as the likes, shares, saves, and comments that a video receives. It also takes into account the user's own engagement on the platform, including who they follow and what videos they interact with. By analyzing this data, TikTok's algorithm can identify patterns and preferences, allowing it to recommend videos that are more likely to resonate with each individual user.
  - Moreover, TikTok's data-driven approach likely extends beyond user engagement. The algorithm may also consider key information associated with each video, such as keywords, captions, hashtags, and any embedded details. This additional data helps enhance the accuracy of the recommendations provided by the platform.
- Is data and analytics viewed as a "*profit center*" or a "*cost center*"?
  - At TikTok, data and analytics are primarily viewed as a "profit center." This is because a significant portion of TikTok's revenue is generated through advertising. The ability to precisely target advertisements based on user data enhances the effectiveness of TikTok's advertising platform, which in turn drives its profits. Advertisers are willing to invest in TikTok's advertising services because they can reach their desired audience with greater precision, resulting in a higher likelihood of conversions and sales. Therefore, at TikTok, data analytics plays a crucial role in driving revenue and is considered a "profit center", enabling TikTok to optimize its advertising capabilities and deliver value both to its users and advertisers.
  -

Company Introduction
- informative description of the field or company, including popular lines of business.
- That is, how do they make money? What do they sell?
    - Short video - Paid Promotion
    - E-commerce - Commission
    - Advertising - Feed Ads

Data and Analytics in Business Model
- How are analytics and data used in the business model?
    - To answer this question, **you should pick a <u>research paper</u> or a <u>blog post</u> from a company's tech blog** where they walk through a specific business application using analytics.
    - Your presentation of this paper/blog should be specific and technical.
    - You may be asked technical questions, so be sure you know the content!
    - The less vague the better.

    - Attribution Analytics. The first feature, included as part of this release, is Performance Comparison, a measurement tool that visualizes conversions across different time windows to help advertisers find an attribution strategy that works for their business.
    - A conversion is any user action that has business value e.g. a user purchasing from a site or advertisement. Either that or they subscribed.
    - first-party measurement solution to enable web advertisers to go beyond the last-click model and truly understand the customer journey
    - Not all conversions are received through common methods. 79% of all conversions attributed to TikTok by users were missing from last-click models.
    - Limitations of Static Models for TikTok's Recommendation System
    - To be a good TikTok recommendation system, it has to be scalable, flexible, adept at handling non-stationary data and can keep learning from extensive user data provided. Later we will go through why today's TikTok recommendation system, monolith, has all those traits. All those traits are pretty attractive but non-stationarity/the ability to extract key information from data to improve itself is quite important to TikTok because user preference is always changing.
    - That's why we are going to briefly go through why certain static models are not ideal for TikTok before we get into Monolith.
    - Starting with SQL databases, they struggle with TikTok's rapid data flow and managing the platform's diverse, unstructured data like videos and user comments. This mismatch limits their effectiveness in handling TikTok's massive data volume.
    - Next, linear regression models. These models fall short in understanding the nonlinear, intricate relationships between user behaviors and content preferences on TikTok. This linear approach doesn't match with the complex dynamics of user interaction on the platform.
    - The last one is decision trees. While adding complexity, that is adding more judging criteria or branches, seems like a solution, it actually leads to overfitting,

- where models become too specific to training data and lose their predictive power on new, diverse user data. In addition, these complex trees can slow down the real-time responsiveness crucial for TikTok.
  - In summary, each of these models, while effective in certain scenarios, doesn't fully meet the requirements of a dynamic, data-rich environment like TikTok. That's why we need a more sophisticated model like Monolith. With that, I will pass it onto Leo.
- leo
  - Monolith: collisionless embedding table with optimizations such as expirable embeddings and frequency filtering to reduce its memory footprint; second, it provides a production-ready online training architecture with high fault-tolerance; finally, it proved that system reliability could be traded-off for real-time learning. Monolith has successfully landed in the BytePlus Recommend product.
  - In pursuit of a better customer experience, delivering personalized content for each individual user as real-time response is a common goal of these business applications.
  - Despite the ubiquitous adoption of production-scale deep learning frameworks e.g. TensorFlow or PyTorch, these general-purpose frameworks fall short of business demands in recommendation scenarios for various reasons: on one hand, tweaking systems based on static parameters and dense computations for recommendation with dynamic and sparse features is detrimental to model quality; on the other hand, such frameworks are designed with batch-training stage and serving stage completely separated, preventing the model from interacting with customer feedback in real-time

https://arxiv.org/pdf/2209.07663.pdf

Job Titles and Descriptions
- What are some job titles and descriptions of different analytics positions?
  - Do they distinguish between data analysts, data scientists, and data engineers? Yes
  - Some examples: Big Data Engineer, Data Solutions Specialist/Architect, Data Analyst, Data Scientist, etc
- What are the average job qualifications?
  - At least a bachelor's in some technical field, especially engineering, software, etc.
  - Experience in some coding languages, which vary depending on the position, but including Python, Java, C++, Swift, Objective C, JavaScript, etc

- - Other than these basics, the qualifications are intentionally broad, referencing things like "strong understanding of algorithms" or "good communication", etc.
    - Job specific preferred requirements, such as previous experience in an internship/research in that particular field
    - For most positions, submitting projects or portfolios may not be explicitly required, but is almost certainly an unspoken requirement
  - What can you expect from a job at this company or in this field?
    - Work-life tradeoff: All work, relentless pursuit of productivity, balance nonexistent
    - Salary expectations (a good source is the [the H1-B database](), [levels.fyi](), Blind, etc.) Median salary 205000, however even for lower level/entry level position you can expect six figures
    - Time off or family leave: 10 paid holidays per year plus 17 days of Paid Personal Time Off (PPTO) (prorated upon hire and increased by tenure) and 10 paid sick days per year as well as 12 weeks of paid Parental leave and 8 weeks of paid Supplemental Disability.
    - Retirement packages (401k matching, pensions, etc.) 401k matches 5% up to your 10%, offers stock options
    - Base versus bonus: Year end bonuses in addition to typical six figure+ salary
    - Company culture (for example [Amazon doesn't have the best track record]()) Nonexistent work life balance, employees commonly leave after short time, burnout is common, poor communication between management and employees, company practices are largely influenced by parent company ByteDance, which creates a much more chinese style culture emphasizing work
    - We cover 100% premium coverage for employee medical insurance, approximately 75% premium coverage for dependents and offer a Health Savings Account(HSA) with a company match. As well as Dental, Vision, Short/Long term Disability, Basic Life, Voluntary Life and AD&D insurance plans. In addition to Flexible Spending Account(FSA) Options like Health Care, Limited Purpose and Dependent Care.
  - Are there any examples of the interview process?
    - Questions vary greatly depending on the position, but you can expect a few rounds of coding and behavioral interviews
    - Examples: Leetcode questions, questions about languages you list on your resume

- - - ● Behavioral question examples: What are your strengths and weaknesses? How do you prioritize tasks? Describe a situation that you need to dig out the cause? Why do you want to work for Tiktok?
      - ● They do encourage using the STAR interview method
    - ■ What was your overall takeaway from this research process?
      - ● Did it make you want to go further in this field? Hell no(just me tho)
      - ● Did it make you want to go into another field?
      - ● Or did it make you want a completely different life direction? Yes

- ○ There should be **four** questions in total.
  - ■ The **first** question should be a basic Python coding challenge. This question should be based on the tools from the Intro to Python sections of the course.
  - ■ The **second** question should be a data wrangling question with a dataset. This question should be based on the tools from the Pandas sections of the course.
    - ● The dataset will ideally be relevant to the company or industry, but I do realize this is a hard task, so you can use a toy dataset or one of the datasets used in class.
  - ■ The **third** question should be a machine learning regression or classification problem based on the dataset from the second question. This question should be based on the tools from the Scikit-Learn sections of the course.
  - ■ The **fourth** question should be a behavioral question and you do **not** need to provide a solution to this question.
- ○ The first three questions should include code for solutions and any relevant explanations.
- ○ Feel free to use sample interview questions as an inspiration, but you should have your own twist on the questions. Nothing should be directly copied.

Interview Question 1: Write a Python function that takes a list of integers as input and returns a new list containing only the even numbers from the original list.

Solution 1:
```
def get_even_numbers(input_list):
    even_numbers = []

    for num in input_list:
        if num % 2 == 0:
```

```
        even_numbers.append(num)

    return even_numbers

numbers = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]
result = get_even_numbers(numbers)
print(result)
```

Interview Question 2: Given a dataset in CSV format that contains information about employees in a company, including columns for employee ID, name, department, salary, and years of experience. Write a Python function that reads the CSV file and returns the average salary of employees in each department.

Solution 2:
```
import pandas as pd

def calculate_average_salary(csv_file):
    df = pd.read_csv(csv_file)

    # Group the data by department and calculate the average salary
    avg_salary_by_dept = df.groupby('department')['salary'].mean()

    return avg_salary_by_dept

csv_file = 'employees.csv'
result = calculate_average_salary(csv_file)
print(result)
```

Interview Question 3: You are working on a regression problem and want to build a machine learning model using Scikit-learn. The task is to predict the housing prices based on various features such as the number of bedrooms, square footage, and location. Write a Python code snippet that demonstrates how to train a regression model using Scikit-learn's linear regression algorithm and make predictions on new data, assuming you have a dataset with features X and target variables Y

Solution 3:
```
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error

import numpy as np
import pandas as pd

np.random.seed(42)
```

```python
# Generate random features (X)
num_samples = 1000
num_features = 3

bedrooms = np.random.randint(1, 5, size=num_samples)
square_footage = np.random.randint(500, 3000, size=num_samples)
location = np.random.choice(['A', 'B', 'C'], size=num_samples)

# Generate random target variable (y) based on features
price = 100000 + 50000 * bedrooms + 200 * square_footage + np.random.normal(0, 10000,
size=num_samples)

# Create a DataFrame
data = pd.DataFrame({'<link>Bedrooms</link>': bedrooms, '<link>Square Footage</link>':
square_footage, '<link>Location</link>': location, '<link>Price</link>': price})

# Save the dataset to a CSV file
data.to_csv('housing_data.csv', index=False)




# Split the dataset into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

model = <link>LinearRegression</link>()

# Train the model using the training data
model.fit(X_train, y_train)

# Make predictions on the test data
y_pred = model.predict(X_test)

mse = <link>mean_squared_error</link>(y_test, y_pred)
print("Mean Squared Error:", mse)
```

Interview Question 4: How do you handle high pressure situations?