

**avito.tech**

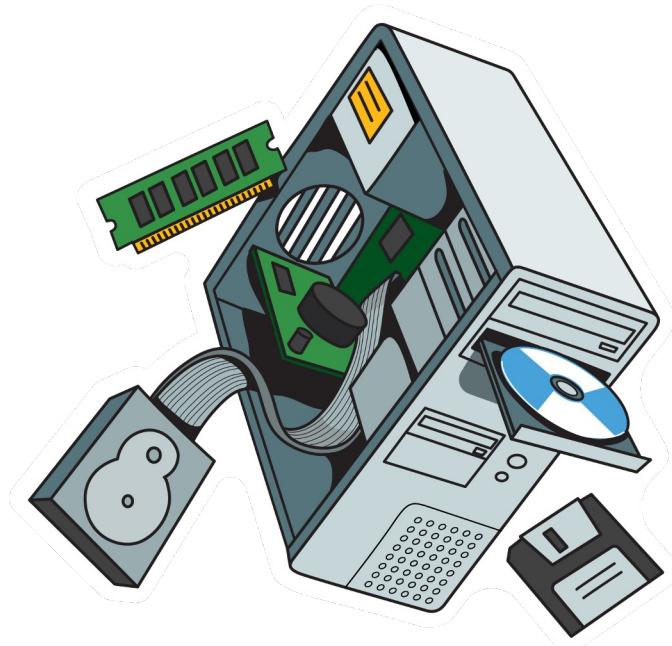
# FORECASTING FOR MEDIA MARKETING EVALUATION

**ARKADIY VORONKIN**  
Senior Data Analyst in Marketing Team



# AGENDA

- 1. Media Marketing**
  - Channels
  - Specifics
- 2. Forecasting techniques in marketing**
  - How do you use it?
  - Pros & Cons
- 3. Marketing Custom Forecast Tool**
  - Goals of the Tool
  - Covariates
  - FBProphet
  - Cross-validation, error calculation
  - Interpretation
- 4. Instrument testing**
  - MAPE vs MTAPE comparison
- 5. Further steps**



# Arkadiy Voronkin

Senior Analyst in Marketing Team

Bachelor in Economics and Finance



Masters in Data Science and  
Business Analytics



Analyst in Avito (~ Craigslist; Ebay)



Streams I lead in Marketing Analytics:

- Media Marketing
- Web Analytics
- SMM & PR



# MEDIA MARKETING

WHAT'S THIS?  
WHY IS IT SPECIAL?



# MEDIA MARKETING

## DEFINITION



General Message: Tells about a product **in general** or about **new features** of the product



Channels with **maximum coverage** are used



DoubleClick  
Bid Manager



Online  
**(can** track cookie id)



Offline  
**(can't** track cookie id)

# MEDIA MARKETING

## SPECIFICS

Offline channels aren't tracked on **user-based level => we cannot attribute payments** to users' exposure to marketing



**Standard** marketing instruments are **not available** for offline media



Main goal is to **estimate the efficiency** of offline media (ROI) for improving the **media budget split**

# FORECASTS FOR MEDIA EVALUATION



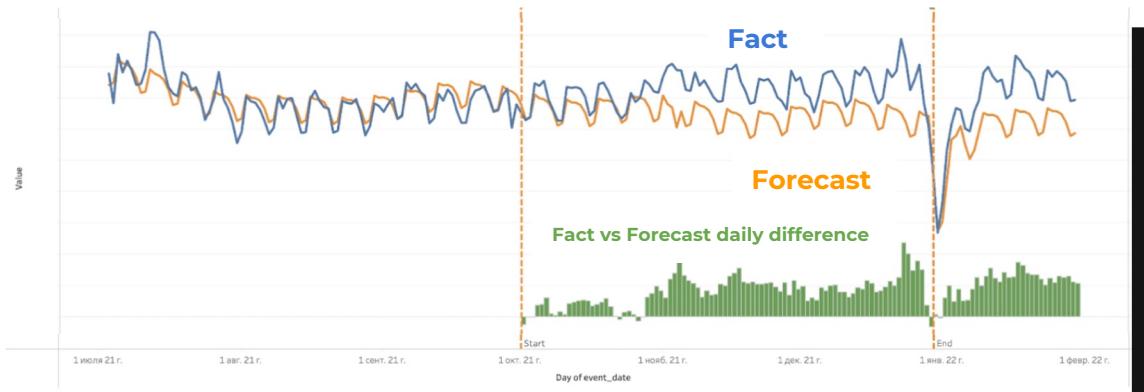
# FORECASTS FOR MEDIA EVALUATION

The forecasts can answer the following question:

**How would have the metric behaved if we had not launched the campaign?**

Main components of forecasts:

- **Forecasts** is built as if **no campaign is launched**
- **Fact** is built with regard to the **campaign effect**



# PROS OF FORECASTS



We can include some covariates (e.g. control regions) in the model controlling for some external factors =>  
**we can estimate both federal and regions campaigns**



Carefully decomposing the time series into trend and seasonality we can achieve  
**high accuracy of the estimation**



Some ready-to-use libraries like Causal Impact and FBProphet  
• make it much **easier and faster to build a forecast**

# CONS OF FORECASTS



Highly-volatile metrics with no clear seasonality can yield in **high forecast** error leading to **low power** of the instrument



Impossible to **distinguish between the specific channels** in marketing. Moreover, impossible to diminish the exposure to **side-effects**



The effect/uplift is **diminishing for regular campaigns**, i.e. those

- that are repeated over the years

# MARKETING CUSTOM FORECAST TOOL



# INSTRUMENT GOALS

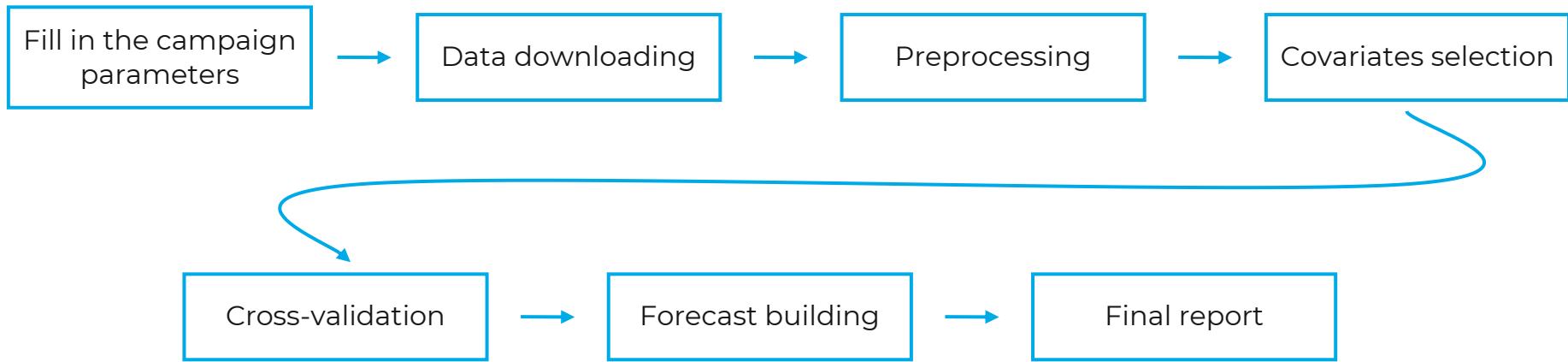


What did we aim at specifically?

- ▶ To develop the **powerful instrument with low forecasts error**
- ▶ To be able to evaluate the effect on **every metrics and every dimension** (region, category, etc.)
- ▶ To make estimation **as fast as possible** (1 day max instead of several weeks)
- ▶ To make the **self-sufficient instrument** that would not require constant analytical support
- ▶ To make the instrument **user-friendly and available for business managers** to use

# MARKETING CFT

## AUTOMATED PIPELINE



# MARKETING CFT

## CAMPAIGN PARAMETERS AND DATA DOWNLOADING

Managers fill in a simple Google Sheet table with the campaign parameters

```
def req_sql_builder(
    target_region,
    metric_name,
    vertical,
    log_cat,
    excluded_locations
):
    sql = """select
        event_date as ds
    """
    if log_cat == 'target':
        sql = sql + f",case when locationname in ('{target_region}' \n+ 'and') then 'target' else locationname end as region"
    sql += f"""
        ,sum(metric_value) as y
        from DMA_Regional_Db_Metrics ab
        join DMA_Current_Locations cl using (location_id)
        where user_segment_market = 'any'
        and is_user_asd = 'any'
        and is_national = 'any'
        and cl.City_Population_Group != 'under_100k'
        and platform_id = -1
        and event_date=date_trunc('year',TIMESTAMPADD(year,-3,CURRENT_TIMESTAMP)):date
    """
    sql = sql +
    f"+ where vertical_id in (select distinct vertical_id from DMA_Current_Microcategories where vertical = {vertical} \n+ 'and') "
    if log_cat != "Any":
        sql = sql +
        f"""
        ,log_cat as l
        ,and logical_category_id in (select distinct logical_category_id from DMA_Current_Logical_Categories where logical_category = {log_cat} \n+ 'and' \n+ log_cat + ''))
        """
    else:
        sql = sql + "\n        ,and logical_category_id = -1"
    sql = sql +
    f"""
        ,metric_name as m
        ,sql = sql +
        f"""
        ,\n        and ab.location_id not in (select location_id from DMA_Current_Locations where isad = 'no'
        + 'and' + target_region + 'and' + ab.location_id not in (select location_id from DMA_Current_Locations where isad = 'yes'
        + ')') and locationname not in ('{target_region}' \n+ ')') and ParentLocationId=2885)
        """
    sql = sql +
    f"""
        ,\n        and cl.parentlocation_id not in (select location_id from DMA_Current_Locations where
        + target_region + ',' + excluded_locs + '))
        """
    if 'CAKWR-NereipyD' \n+ in excluded_locations:
        sql += sql +
    """

```

Through the **API** parameters are uploaded in **SQL builder** which download the data in Python

**Fill in the campaign parameters**

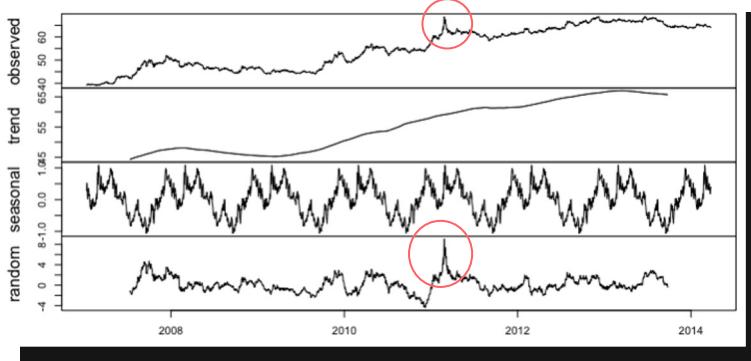
**Data downloading**

**Preprocessing**

# MARKETING CFT

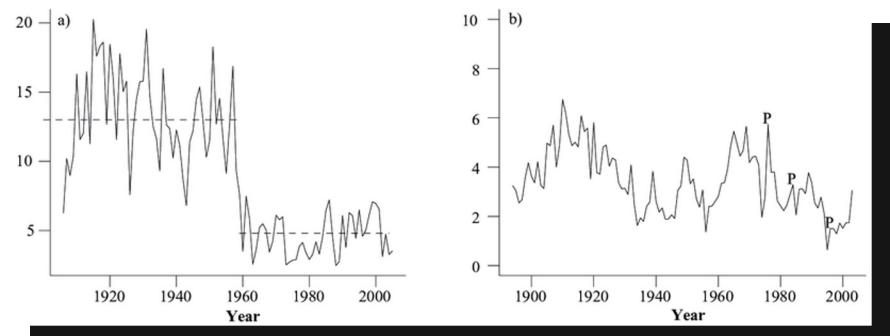
## PREPROCESSING

Outliers detection with interpolation



$$x_i \subset \text{outliers} \iff x_i \subset \pm 3 * IQR_x$$

Changepoints detection



Fill in the campaign parameters



Data downloading



**Preprocessing**



# MARKETING CFT

## COVARIATES SELECTION

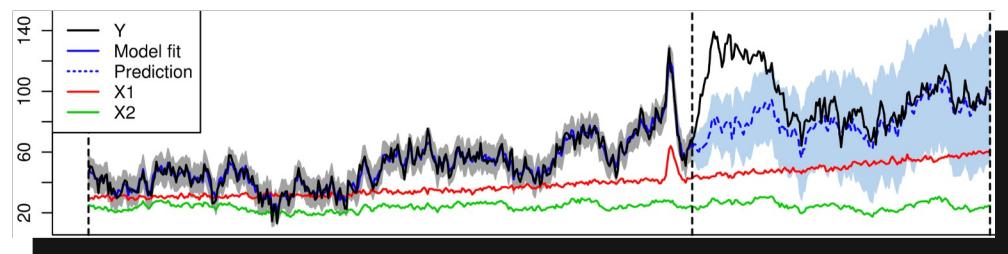
**Covariate** - exogenous factor that correlates with target metric and thus improves the forecast

Covariate criteria:

- ▶ high correlation with target metrics
- ▶ **no correlation with marketing activity**

Good covariates example:

- ▶ Regions with no marketing activity in them
- ▶ Economics factor (key rate, exchange rate)
- ▶ Supply metrics (number of available SKUs)  
(only for buyer's campaigns!)



**Covariates selection**



Cross-validation



Forecast building



Final report

# MARKETING CFT

## COVARIATES SELECTION. EXAMPLE.

Target metrics: number of deliveries demanded

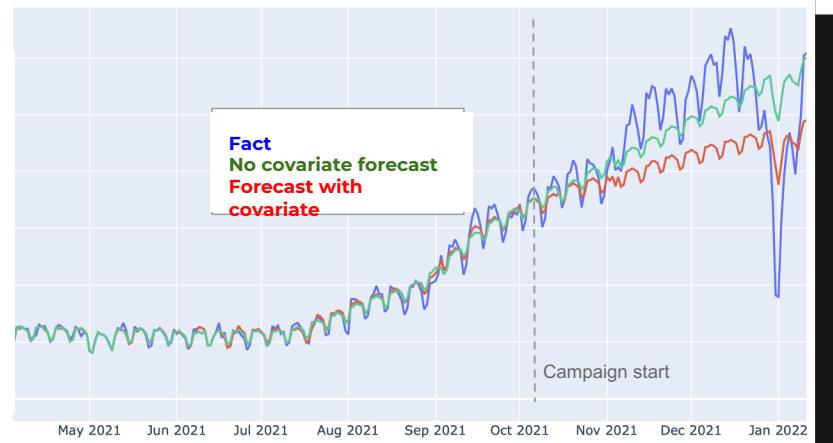
Covariates: average delivery price

Problem:

- numerous changes in delivery price do not allow to distinguish between **the price effect and the marketing effect**
- due to its instability (new feature) the **forecasting error was 17% (MAPE)**

Solution:

Using average delivery price as a covariate, we controlled for price effect, isolating it from marketing one and obtained an **error of only 8%**



Covariates selection



Cross-validation



Forecast building



Final report

# FBPROPHET FORECAST

$$y(t) = g(t) + h(t) + s(t) + e(t)$$

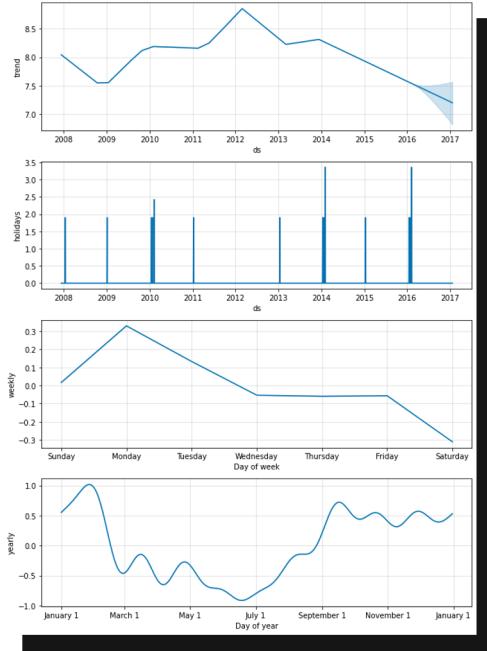
**y(t)** - target (forecasted) metrics

**g(t)** - trend

**h(t)** - holidays (anomaly periods)

**s(t)** - seasonality (weekly, monthly, yearly)

**e(t)** - random component



Covariates selection



**Cross-validation**



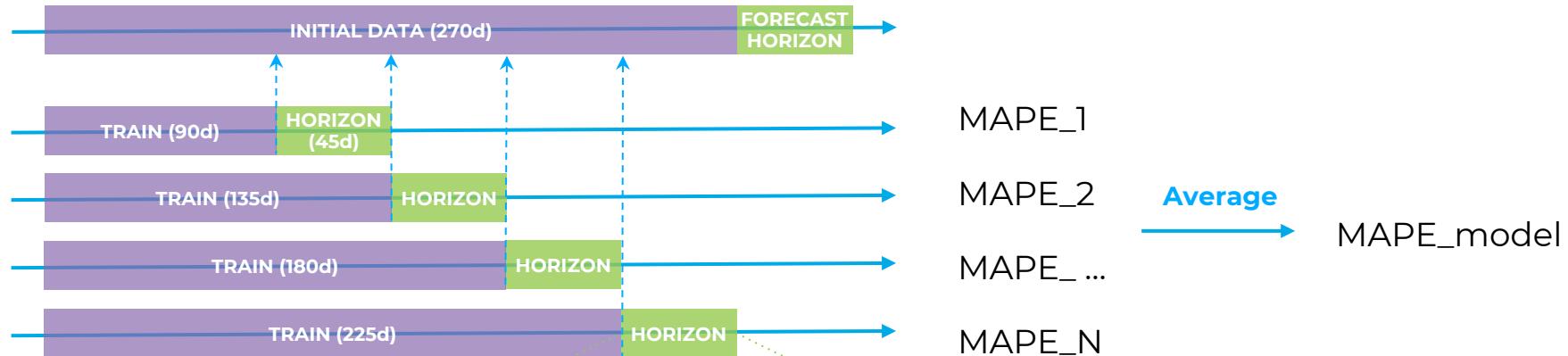
Forecast building



Final report

# MARKETING CFT

## PARAMETERS GRID SEARCH. (EXPANDING WINDOW CROSS-VALIDATION)



$$MAPE = \frac{1}{H} \sum_{t=1}^H \left| \frac{fact_t - forecast_t}{fact_t} \right|$$



# MARKETING CFT

## PARAMETERS GRID SEARCH. (EXPANDING WINDOW CROSS-VALIDATION)

Model 1	param1=0.01, param2=0.5, param3 = 'additive', ...	MAPE_model = 0.064
<b>Model 2</b>	<b>param1=0.1, param2=0.25, param3 = 'additive', ...</b>	<b>MAPE_model = 0.056</b>
Model 3	param1=0.05, param2=0.1, param3 = 'multiplicative', ...	MAPE_model = 0.089
Model 4	param1=0.25, param2=0.01, param3 = 'multiplicative', ...	MAPE_model = 0.070

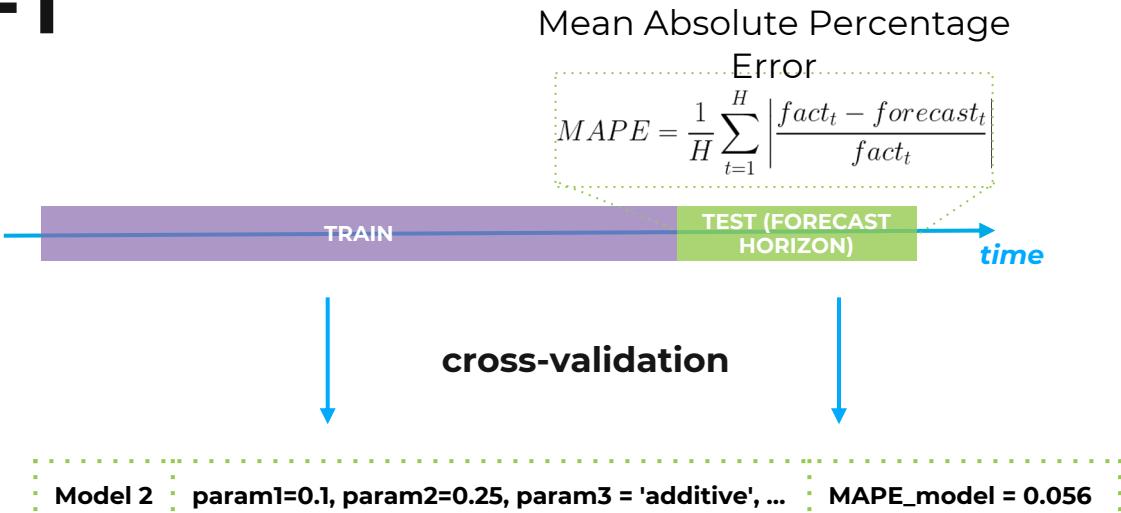
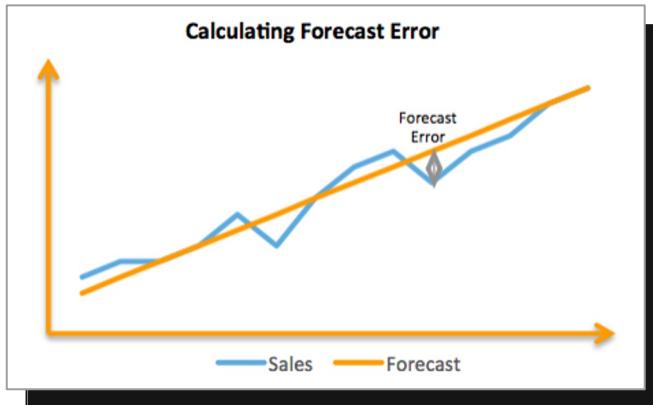
Parameters, **minimizing** the average model error



# MARKETING CFT

## FORECAST ERROR ESTIMATION

Calculating Forecast Error



Parameters, minimizing the average model error

Covariates selection

Cross-validation

Forecast building

Final report

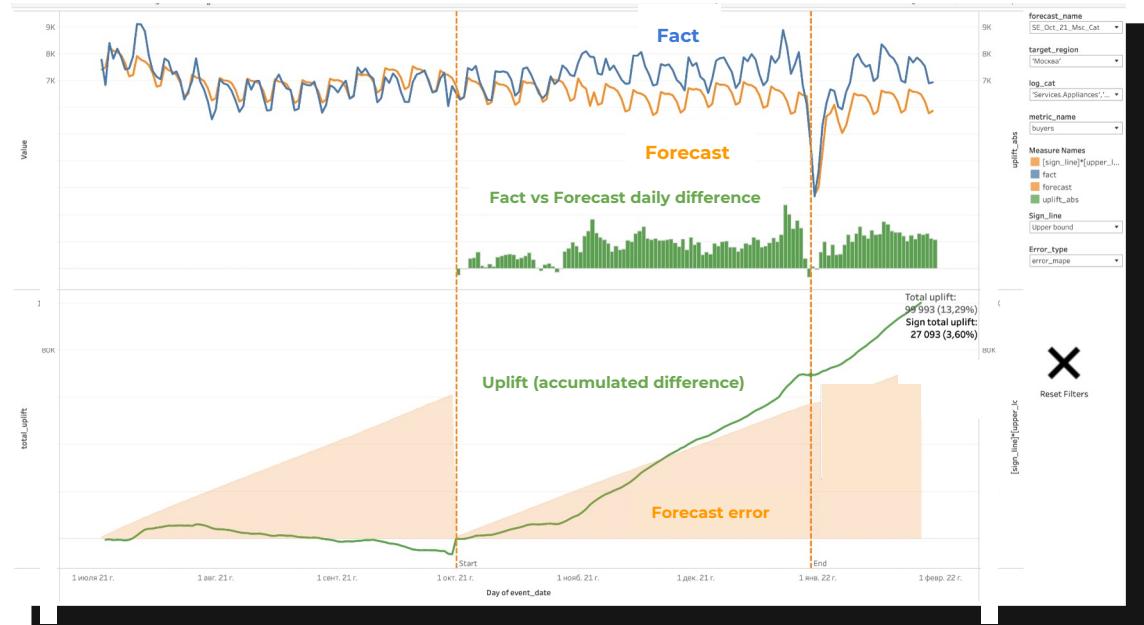
# FORECASTS FOR MEDIA EVALUATION

The forecasts can answer the following question:

**How would have the metric behaved if we had not launched the campaign?**

Main components of forecasts:

- **Forecasts** is built as if **no campaign is launched**
- **Fact** is built with regard to the **campaign effect**



Covariates selection



Cross-validation



**Forecast building**



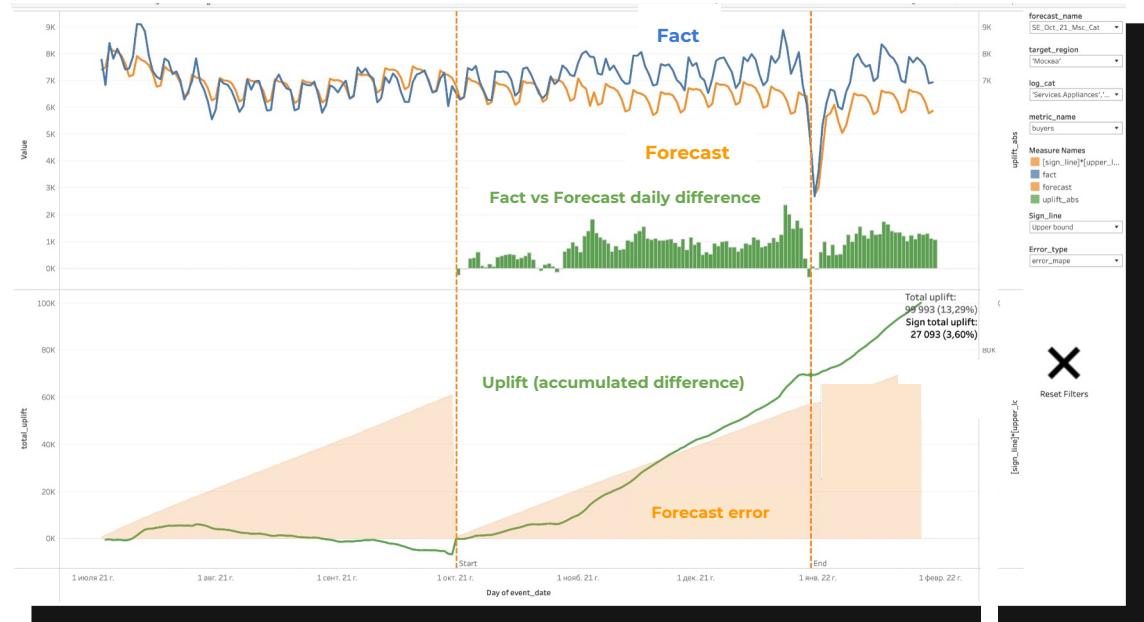
Final report

# FORECASTS FOR MEDIA EVALUATION

**Accumulated difference (uplift)** by the end of the campaign - **total marketing effect**

**Forecast error** - average difference between the fact and forecast **before the campaign**

**Statistical significance criterion:**  
**uplift > forecast error**  
by the end of the campaign



Covariates selection



Cross-validation



**Forecast building**



Final report

# MARKETING CFT 2.0

## REST OF AUTOMATIZATION



Single **python script** is regularly activated via [cron](#) and **Tableau report** is updated on **daily basis**

Covariates selection



Cross-validation



Forecast building



**Final report**

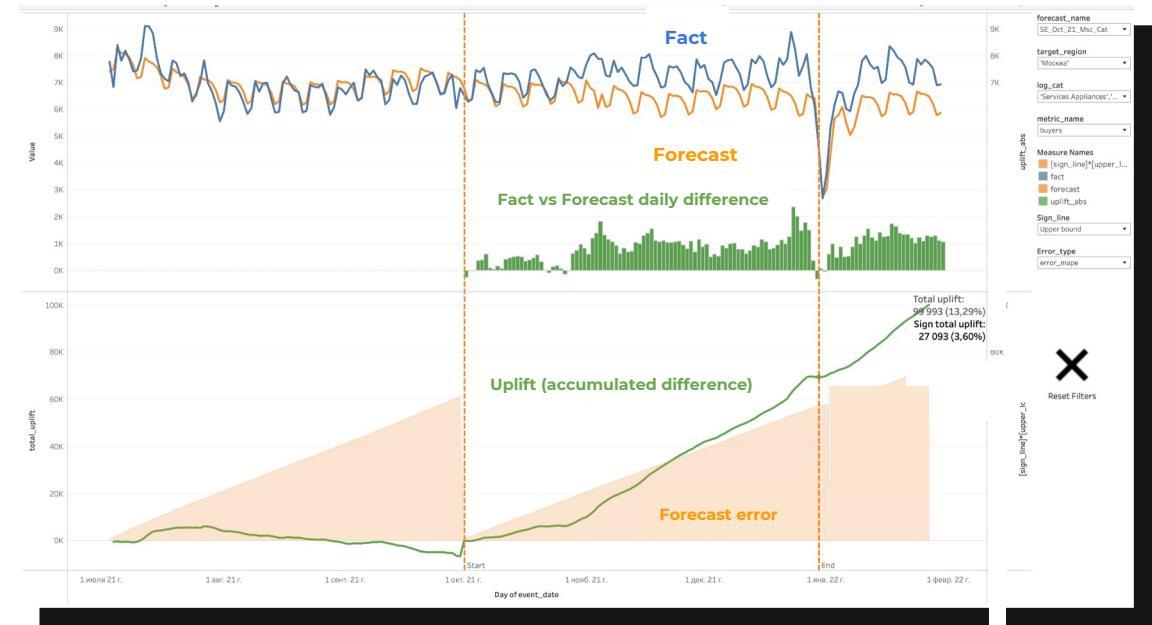
# INSTRUMENT TESTING



# MARKETING CFT

## STAT SIGNIFICANCE CRITERION

Statistical significance criterion:  
**uplift > forecast error**  
by the end of the campaign



# MARKETING CFT

## FORECAST ERROR CHOICE



$$MTAPE = \left| \frac{\sum_{t=1}^H fact_t - \sum_{t=1}^H forecast_t}{\sum_{t=1}^H fact_t} \right|$$

Error calculated for the **whole forecast horizon**. Marketing goal - estimate the effect **during the whole campaign period**.



$$MAPE = \frac{1}{H} \sum_{t=1}^H \left| \frac{fact_t - forecast_t}{fact_t} \right|$$

Average error for **each day** during the forecast period.

**Statistical significance criterion:** by the end of the campaign **uplift > forecast error ↓**

# MARKETING CFT

## CRITERION TESTING. TEST SETUP.

Dataset: ~ **420 time series of revenue** in various regional, categorical dimensions.  
Each time series is 21 day long.

Necessary condition for each series: **no marketing activities present**



# MARKETING CFT

## CRITERION TESTING. TEST RESULT.

	Type I error (AA Test)	Power (AB Test)	Result
Forecast with <b>MAPE</b>	0.0394	0.6919	Low power
Forecast with <b>MTAPE</b>	0.4339	0.8863	High type I error

Decision: use **MAPE** with lower type I error, but aim to **improve the instrument** for higher power.



# MARKETING CFT

## RESULTS



Can estimate the effect of each end every marketing campaign (any metrics, any dimension)

Use covariates for regional and federal campaigns allowing for **lower forecast error**

Instrument is fully automated



**Need to improve the power of the instrument**

# MARKETING CUSTOM FORECAST TOOL

FURTHER STEPS



# MARKETING CFT

## FURTHER STEPS

- 01.** Preprocessing & models ensembles in [new library Kats](#)
- 02.** New stat. criteria testing
- 03.** Search for new covariates



# avito.tech

# Thanks!

# Time for Q&A

funny quotes about forecast

Google search results for "funny quotes about forecast". The top result is a link to "HILLBILLY WEATHER REPORT" with an image of a man in overalls standing next to a world map. Below it is a link to "TODAY'S FORECAST" with an image of a sunset over water. To the right is a box titled "Facebook Weather Forecast" with text about expecting drama and bullsh!t blowing from all directions. Another box says "WEATHER FORECAST FOR TONIGHT: DARK." with a quote by George Carlin.

A grid of 10 cards. Top row: 1. "I miss hating the summer heat." (someecards) 2. "Fool me once, shame on you. Fool me twice, shame on me. Fool me 92,748 times, you are a weather man." (pinterest.com) 3. "WEATHER as told by Snoop Dogg" (quotemaster.org). Bottom row: 1. "Play in the puddles!" (vectorstock.com) 2. "Well it's official, the weather has been diagnosed with bipolar disorder!" (quotesgram.com)

A grid of 10 cards. Top row: 1. "IF YOU HAVE TO FORECAST, FORECAST OFTEN." (picturequotes.com) 2. "Dear old man winter, I think you need a vacation you have worked really hard this year. It's time to let spring start pulling some weight around here too! Sincerely, Enough Winter" (quotemaster.org) 3. "MINNESOTA WEATHER FORECAST" (quotemaster.org) 4. "TODAY'S FORECAST: 99% CHANCE OF WINE." (picturequotes.com) 5. "The goal of forecasting is not to predict the future but to tell you what you need to know to take meaningful action in the present" (azquotes.com) Bottom row: 1. "If you have to forecast, forecast ..." (picturequotes.com) 2. "Quotes about Funny weather (24 quotes)" (quotemaster.org) 3. "Funny Weather Quotes ..." (quotemaster.org) 4. "99% chance of wine | Picture Quotes" (picturequotes.com) 5. "TOP 25 FORECASTING QUOTES | A-Z Quotes" (azquotes.com)

A grid of 10 cards. Top row: 1. "Uber Humor" (steve holl) 2. "WEATHER FORECAST FOR TONIGHT: DARK, CONTINUED DARK OVERNIGHT, WITH WIDELY SCATTERED LIGHT BY MORNING." (quotemaster.org) 3. "Weather forecast for tonight: Dark with a chance of tomorrow in the morning" (coolnsmart.com) 4. "Actually YES! the weather IS great up here!" (yourecards.com) 5. "It's so cold outside I just saw a teenage boy with his pants pulled up." (quotemaster.org) Bottom row: 1. "Funny weather ... ro.pinterest.com" 2. "Forecasts Quotes. Quo... quotesgram.com" 3. "Weather forecast for tonight: Dark ... coolnsmart.com" 4. "Quotes about Funny weather ... quotemaster.org" 5. "Funny Weather Quotes And Sayings ..." (quotesgram.com)